HERIOT-WATT UNIVERSITY

DOCTORAL THESIS

Child Human Capital in Developing Countries

Wei Fan

Submitted for the degree of Doctor of Philosophy

School of Social Sciences

Department of Accounting, Economics and Finance

December 20, 2019

Abstract

This thesis contributes to three current topics in child human capital development in the developing world. First, we examine the role that child ability plays in parental investment decisions in Ethiopia. Second, using a different dataset, we investigate whether women's empowerment could improve child nutritional status in Ethiopia. Third, we study the effect of sanitation on child cognitive ability in India, and explore heterogeneous effects by child endowments.

In Chapter 2, we present a paper which exploits the longitudinal Young Lives data survey in Ethiopia to evaluate the causal effect of child cognitive skill on parental educational investment within the family. The study uses instrument variables combined with sibling fixed-effects to tackle the endogeneity in child ability and parental investment. We find that parents compensate the low-ability child among their primary school-age offspring through increased educational fees. We also find that this effect mainly holds for the high socioeconomic households.

Chapter 3 investigates the association between maternal autonomy and child nutritional status among children aged 0-59 months in Ethiopia. Using the nationally representative Demographic and Health Survey data and an innovative empirical methodology - 'post-double-selection Lasso' (PDS-LASSO) - which helps to avoid dubious variable selection and to deal with omitted variable bias, we provide evidence that a child is less likely to be underweight if his/her mother has high autonomy. Notably, this correlation is strongest for children who are older than two years old.

The last paper, shown in Chapter 4, explores the impact of shocks and policy during the fetal and infancy periods, along with their interaction, on later cognitive performance in rural India. Specifically, it studies the separate effect of in-utero rainfall fluctuations and a sanitation campaign at birth. It further discusses whether the effect of the sanitation campaign is differential by child endowments induced by previous rainfall shock. Using a difference-in-differences design, we do not find that the return to the sanitation programme is higher for advantaged children, as evidence shows that children who were exposed to positive rainfall in utero achieve similar scores at the age of 8-11 through the improved sanitary environment at birth.

Acknowledgements

First and foremost, I am most grateful to have Cath as my first supervisor. I am fortunate to have her company along the journey of my PhD. She has been the light in my research process, academically and personally. She offers me structural solutions when I face obstacles in research and motivates me by her good African cooking in the 'lovely' Scottish weather. Apart from her instrumental help in my thesis, she generously offers lots of opportunities in work, presentations, and internship, which turn out to be significant assets for my career. Her care and support for colleagues also inspire me a lot. I also like to thank Philippe, my second supervisor, for being patient, supportive and encouraging in my career.

Thanks to the lecturers, such as Mark, Andrea, Atanas, Shumei, and Arnab in the department for attending my presentations and giving valuable advice. Thanks to my economics PhD cohorts and all those who joined the 12pm lunch and played in our football team: Rong, Jan, Antonio, Fran, John, Ryuta, Achim, Yeshwas, Sering and Romi. It is a great pleasure to study and work with you. Thanks to the PhD office mates in the Department of Accounting, Finance and Economics: Nana, Antonio, and Dara. I am appreciative of Vincenzo and the rest of the team in SPESI, UNICEF Ethiopia for welcoming me with the warmest heart and working with me.

Thanks to my family in Buccleuch Free Church. Their love and friendship have been the most supportive solid rock for my life in Edinburgh. Special thanks to my dear friends Shen, Yuanyuan, Moji, Cia, Moira, Eric, Anne, Willie, Daniel, Marion, Joan, Donald, James, Vickie, Heather, Bob, Neil, Mary, Roddy, Christine, Ena, William, Loren, the Blackford Hill home group, and the EDGE teens group. My European friends and friends back in China have been my most excellent support and laughter boosters. Thanks to Yanni, Sarah, Apple, Shiyao, Karen, Jerry, Mara, Marilou, Zhihai, Kara, Class Two in Beijiang Middle School, and 101.5 Dorm at GDUFS.

Last but not least, I would like to express my gratitude to my parents, grandparents, uncles, aunts and cousins in China for their unconditional love and support when I am far away from them in the past five years. I want to express my special gratitude to my mom, who has been a great inspiration and role model in achieving excellent performance in education and career as a female. Thank you for all the efforts you have put in me and supported me in so many ways. Thank you for encouraging me to pursue what I want when I lack confidence.

Joint Work Declaration

Chapter 2 is a joint paper with Dr. Catherine Porter, with my contribution on developing its idea, data cleaning, estimation method, results generation, and writing.

This paper has been accepted by the Journal of Population Economics.

Research Thesis Submission

Name:	Wei Fan		
School:	School of Social Science		
Version: *(i.e. First, Resubmission, Final)*	Final	Degree Sought:	Doctor of Philosophy, Economics

Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

1. The thesis embodies the results of my own work and has been composed by myself
2. Where appropriate, I have made acknowledgement of the work of others
3. The thesis is the correct version for submission and is the same version as any electronic versions submitted*.
4. My thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
5. I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.
6. I confirm that the thesis has been verified against plagiarism via an approved plagiarism detection application e.g. Turnitin.

ONLY for submissions including published works

7. Where the thesis contains published outputs under Regulation 6 (9.1.2) or Regulation 43 (9) these are accompanied by a critical review which accurately describes my contribution to the research and, for multi-author outputs, a signed declaration indicating the contribution of each author (complete)
8. Inclusion of published outputs under Regulation 6 (9.1.2) or Regulation 43 (9) shall not constitute plagiarism.

* *Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.*

Signature of Candidate:		Date:	

Submission

Submitted By *(name in capitals):*	
Signature of Individual Submitting:	
Date Submitted:	

For Completion in the Student Service Centre (SSC)

Limited Access	Requested	Yes		No		Approved	Yes		No
E-thesis Submitted (***mandatory for final theses)***									
Received in the SSC by *(name in capitals):*					Date:				

Inclusion of Published Works

Declaration

This thesis contains one or more multi-author published works. In accordance with Regulation 6 (9.1.2) I hereby declare that the contributions of each author to these publications is as follows:

Citation details	Fan, W. & Porter, C., 'Reinforcement or Compensation? Parental Responses to Children's Revealed Human Capital Levels', Journal of Population Economics, September 2019; working paper 183, Young Lives, 2018
Author 1	Wei Fan has contributed to this paper in developing its idea, data cleaning, estimation method, results generation, and writing.
Author 2	Dr Catherine Porter has contributed to this paper in writing and situating it in the literature.
Signature:	
Date:	

Please included additional citations as required.

Contents

1 Introduction **1**

Bibliography 4

2 Reinforcement or compensation? Heterogeneous parental responses to children's revealed human capital levels in Ethiopia **5**

2.1 Introduction 7

2.2 Literature review 9

2.3 Data and measures 15

2.3.1 Rainfall 16

2.3.2 PPVT scores as a measure of cognitive ability 18

2.3.3 Total educational fees as a measure of cognitive resources 19

2.3.4 Socio-economic status (SES) 23

2.4 Econometric strategy 25

2.5 Results 28

2.5.1 Preliminary results 29

2.5.2 Main results 31

2.5.2.1 IV-FE models: First-stage results and diagnostics 31

2.5.2.2 IV-FE models: Second-stage results 34

2.5.3 Heterogeneity of parental responses to children's early ability 39

2.5.4 Robustness checks 42

2.6 Conclusion 46

Bibliography 48

Appendix 55

3 Is maternal autonomy a potential pathway to child nutritional status? Evidence from Ethiopia 2005-2016 62

3.1 Introduction 64

3.2 Literature review 67

3.2.1 Definition and measurement of women's autonomy . . . 67

3.2.2 The relationship between women's autonomy and child's nutrition 70

3.3 Data 73

3.3.1 Maternal autonomy: main explanatory variable 74

3.3.1.1 Construction of latent autonomy variables: using Bayesian analysis based on a CFA model . 75

3.3.2 Child nutritional status: outcome measures 81

3.3.3 Control variables 82

3.4 Econometric model 85

3.5 Results 89

3.5.1 Preliminary results 89

3.5.2 Main results 92

3.5.3 Heterogeneous results 97

3.5.4 Robustness checks 100

3.6 Conclusion 101

Bibliography 104

Appendix 112

4 Do early life shocks constrain returns to human capital investment? Evidence from a large-scale sanitation programme in India 124

4.1 Introduction 126

4.2 Literature review 130
4.3 Theoretical framework 137
4.4 Data 139
4.4.1 The TSC programme 139
4.4.1.1 Description of programme 139
4.4.2 TSC data 140
4.4.2.1 TSC intensity variable 141
4.4.3 Rainfall data 142
4.4.3.1 Rainfall shock variable 142
4.4.4 IHDS survey data 145
4.4.4.1 Dependent variables 146
4.4.4.2 Control variables 147
4.4.5 Sample 147
4.5 Empirical strategy 150
4.6 Results 153
4.6.1 Effects of rainfall on test outcomes 153
4.6.2 Effects of TSC on test outcomes 156
4.6.3 Effects of rainfall and TSC on test outcomes 159
4.6.4 Robustness checks 164
4.6.4.1 Parallel trends in test outcomes in older children 164
4.6.4.2 Balance of covariates 166
4.7 Conclusion 167
Bibliography 169
Appendix 177

List of Tables

2.1 Summary statistics 24
2.2 Preliminary regression models 30
2.3 First stage regressions: Results and tests of underidentification and weak identification 32
2.4 IV-FE regression models of cognitive ability and total educational fees 35
2.5 IV-FE model of the effects of cognitive ability on total educational fees: potential heterogeneity effect by SES 40
2.6 Robustness regression models 44
A1 Full list of coefficients from Table 4 56
A2 Robustness check: First stage results adding sibling rainfall in infancy using community fixed-effects model 57
A3 Redundancy tests: Cognitive ability and cognitive resources . . 58
A4 Robustness check: First stage results using restricted sample . . 59
A5 IV-FE results: HAZ as the interest variable 60

3.1 Percent of mothers answering autonomy questions in the sample 76
3.2 Sample statistics for estimated factor scores using the Bayesian CFA model 81
3.3 Descriptive statistics for key variables 84
3.4 FE regression models: preliminary results 91

3.5 PDS-LASSO regression models: main results 94
3.6 PDS-LASSO regression models: heterogeneous results 98
C1 Child and parental characteristics 113
C2 Household characteristics 114
C3 PDS-LASSO regression models: variables selected for regressions with dependent variables of being underweight and WAZ 115
C4 PDS-LASSO regression models: variables selected for regressions with dependent variables of being stunted and HAZ . . . 116
C5 Mechanism: heterogeneous effect 117
C6 PDS-LASSO regression models: without unpenalised variables 118
C7 PDS-LASSO regression models: heterogeneous results using restricted sample 119
C8 PDS-LASSO regression models: when women have highest autonomy if they make decision jointly with husband 120
C9 PDS-LASSO regression models: restricted sample to women having only one child under five 121

4.1 Mean of normal rainfall in utero when the cut-off of the definition of normal rainfall varies 144
4.2 Descriptive statistics for key variables 149
4.3 Effects of rainfall on test outcomes 154
4.4 Effects of TSC on test outcomes 157
4.5 Effects of rainfall and TSC on test outcomes 160
4.6 Effects of rainfall and TSC on combined scores 163
4.7 Parallel trends: using cohorts born before TSC's implementation 165
E1 Effects of rainfall on test outcomes using cohorts born between 1999-2004 178

E2 Effects of rainfall on test outcomes using rainfall in the first year of life 179
E3 Effects of TSC on test outcomes using TSC measured by varying population 180
E4 Sanitation-cognition gradient at the age of one and two 181
E5 Balance of covariates: Rainfall shocks, TSC and control variables 182

List of Figures

2.1 Annual rainfall by community, 1998-2008 17

2.2 IRT PPVT scores by language, 2009 19

2.3 Young Lives Survey timings 20

2.4 Total educational fees 21

2.5 Mean cognitive resources and cognitive ability for each fifth percentile of the cognitive ability distribution 22

2.6 Intra-household difference in total educational fees by SES . . . 25

2.7 Kernel density plot of household wealth index by maternal education, 2013 42

B1 Estimated β by direct effect of instrument 61

3.1 Conventional CFA model of autonomy 79

3.2 Trends in nutritional status of children under age five, EDHS 2005-2016 82

D1 Conventional CFA model with standardised factor-loadings . . 122

D2 Bayesian CFA model with standardised factor-loadings, allowing for cross-loadings 123

4.1 Diagram of technology of capability formation 138

4.2 Variation in TSC and rainfall between 2001-2004 150

4.3 An illustration of the design of study 150

Chapter 1

Introduction

The three chapters in this thesis, while each one of them serving as a stand-alone paper, examine the life cycle of human capital formation from in utero to adolescence. Three specific aspects of human capital development are investigated: whether parents respond to the difference in offspring's ability; whether an intangible input - maternal empowerment - contributes to child development; and whether sanitation investments help disadvantaged children catch up.

As noted by Currie and Almond (2011), while most of the "early origins" literature relies on the reduced form estimation, it leaves the question - through which channel, pure biology or responsive investments, the effect is displayed - unanswered. Chapter 2 attempts to answer the question of whether parents compensate or reinforce in the difference in children's revealed ability. Complementing the evidence found in current studies which mainly focus on birth weight as child endowments in the developed context, we look at how parents respond to the revealed difference in primary school-age children's cognitive skill in Ethiopia.

In order to pin down the causal effect, this paper applies an instrumental variable method combined with a sibling fixed-effects to isolate the exogenous variation in child cognitive achievement within the family. Specifically, we exploit rainfall fluctuation during child critical developmental period within the household, which induces an exogenous variation in child cognitive ability observed in later childhood. We find that on average parents tend to invest more in a child with low ability and that higher-class parents compensate more in comparison with lower-class ones.

While early childhood development literature has established particular tangible inputs to child human capital production, such as nutritional intake and home environment, Chapter 3 focuses on a potential intangible input to child nutrition, i.e. women's autonomy. We ask whether a higher level of maternal autonomy is correlated with better child nutrition for children under five in Ethiopia. This paper makes advances to the development literature by adopting a novel methodology, called "post-double-selection Lasso" (Belloni et al., 2014), to deal with variable selection and omitted variable bias, allaying concerns about having too many control variables in the model or dubious '*p*-hacking' selection of variables. We also include village-level fixed effects in the model. This paper uses three rounds of national representative Demographic and Health Survey data between 2005-2016, with rich information on women's autonomy and characteristics of child, parents, and households. We measure mother's autonomy using a reliable statistical method based on the theory of autonomy, i.e. Bayesian Confirmatory Factor Analysis. The results suggest that a child of a mother with a higher level of autonomy is less likely to be underweight, and this is particularly apparent among children aged 24-59 months.

Lastly, in Chapter 4, we look at how sanitation programme would affect child cognition and whether its effect is differentiated by child endowments. In India, more than half of the population defecate openly, exacerbating the disease environment to which vulnerable infants are exposed. In 2001, a national sanitation campaign called Total Sanitation Campaign was introduced, aiming to eliminate open defecation, especially in rural areas. Relying on a difference-in-differences method, we discover the positive impact this sanitation investment at birth makes on cognitive achievement at the age of 8-11 in rural India.

Furthermore, this paper contributes to learning the dynamic aspect of the technology of skill formation. In Cunha et al. (2006), Cunha and Heckman (2007), and Heckman (2006, 2007)'s dynamic human capital formation theory, the authors promote a key feature of skill formation, i.e. "dynamic complementarity", which suggests that the productivity of investment can be driven by the level of skills acquired in the previous stage. To supply empirical evidence, we need to tackle the econometric challenge in finding an exogenous measurement of skill/endowments to be interacting with the sanitation investment. Recalling the context of India, a rainfed country, we use adverse rainfall in utero as an exogenous source of variation in infant endowments. The results suggest that adverse rainfall in utero reduces child test outcomes at the age of 8-11. However, we do not find complementarity that the high level of endowments could raise the return of the sanitation programme. We contribute to the literature in being the first to identify the interaction between sanitation and early endowments and in investigating this effect in the first year of life.

Bibliography

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen (2014). "Inference on treatment effects after selection among high-dimensional controls". *The Review of Economic Studies* 81.2, pp. 608–650.

Cunha, Flavio and James Heckman (2007). "The technology of skill formation". *American Economic Review* 97.2, pp. 31–47.

Cunha, Flavio, James J Heckman, Lance Lochner, and Dimitriy V Masterov (2006). "Interpreting the evidence on life cycle skill formation". *Handbook of the Economics of Education* 1, pp. 697–812.

Currie, Janet and Douglas Almond (2011). "Human capital development before age five". *Handbook of Labor Economics*. Vol. 4. Elsevier, pp. 1315–1486.

Heckman, James J (2006). "Skill formation and the economics of investing in disadvantaged children". *Science* 312.5782, pp. 1900–1902.

— (2007). "The economics, technology, and neuroscience of human capability formation". *Proceedings of the National Academy of Sciences of the United States of America*, pp. 13250–13255.

Chapter 2

Reinforcement or compensation? Heterogeneous parental responses to children's revealed human capital levels in Ethiopia

[1]

[1] This is a joint paper with Dr. Catherine Porter, with my contribution on developing the idea, data cleaning, estimation method, results generation, and writing. This paper has been recently accepted by the *Journal of Population Economics*. It is also a working paper in Young Lives Working Papers Series.

The data are from Young Lives, a 15-year study of the changing nature of childhood poverty in Ethiopia, India, Peru and Vietnam (www.younglives.org.uk). Young Lives is funded by UK aid from the Department for International Development (DFID). The views expressed here are those of the author(s). They are not necessarily those of Young Lives, the University of Oxford, DFID or other funders. Thanks to three anonymous reviewers, Liang Bai, Amalavoyal Chari, Marta Favara, Kalle Hirvonen, Pascal Jaupart, Tatiana Kornienko, Jessica Leight, Patricia Espinoza Revollo, and Mark Schaffer for helpful suggestions, and to seminar and conference participants in Young Lives, Jun 2017; SGPE conference, Jan 2018; CSAE conference, Mar 2018; University of Sussex, Mar 2018; UNICEF Ethiopia, Apr 2018; SEHO conference, May 2018; SMYE, May 2018, and ESPE, June 2018. All errors and omissions are our own.

Abstract

A small but increasing body of literature finds that parents invest in their children unequally. However, the evidence is contradictory, and providing convincing causal evidence of the effect of child ability on parental investment in a low-income context is challenging. This paper examines how parents respond to the differing abilities of primary school-age Ethiopian siblings, using rainfall shocks during the critical developmental period between pregnancy and the first three years of a child's life to isolate exogenous variation in child ability within the household, observed at a later stage than birth. The results show that on average parents attempt to compensate disadvantaged children through increased cognitive investment. The effect is significant, but small in magnitude: parents provide about 3.9% of a standard deviation more in educational fees to the lower-ability child in the observed pair. We provide suggestive evidence that families with educated mothers, smaller household size, and higher wealth compensate with greater cognitive resources for a lower-ability child.

Keywords: Children, Human Capital Formation, Parental Investment, Intrahousehold Resource Allocation.

JEL classification: D13, J1.

2.1 Introduction

A large body of evidence has developed during the past three decades showing that *in utero* and early life conditions have a significant impact on children's early life ability, subsequent development and therefore on outcomes in adulthood (surveyed by Currie and Almond (2011) and Almond et al. (2018)). Most of these studies are reduced-form estimates of the total effect of an early life shock or adverse event on final adult health. However, ability in early life impacts on later human capital not only through the biological channel (Heckman, 2007), but also through the channel of parental involvement - in theory parents can either reinforce or compensate for revealed early ability. It is then an empirical question whether parental actions amplify or mute the ultimate effect of early life shocks and circumstances on adult human capital outcomes.

Our paper contributes to this latter research question, which is of direct policy relevance. The current literature comprises a body of empirical evidence that appears somewhat contradictory, containing studies that document both compensatory and reinforcing behaviour of parents. Attempting to clearly identify such effects given the econometric concerns is extremely difficult, and could be one reason for the apparently conflicting results. Alternatively, there may be important differences across country contexts (either cultural or economic) that are leading to such different conclusions.

Our contribution extends the existing literature in three specific ways. First, we examine the response of parents to differences in child cognitive ability in early childhood in a low-income country, using a measure of ability rather than birth weight or height as a proxy. We are aware of only two previous studies that have analysed parental responses to observed cognitive ability beyond birth. Frijters et al. (2013) find that parents reinforce cognitive resources in response to differences in cognitive ability in the USA. Ayalew (2005) also finds reinforcing effects, but these results are based on estimates from only one village in Ethiopia.[2]

Second, we use both sibling fixed-effects and a plausibly exogenous source instrument (rainfall in early life) for variation in cognitive ability to more convincingly identify parental responses, rather than relying on within-twin estimation, since twins are not the ideal group on which to study such effects (Bhalotra and Clarke, 2018). Other instruments have been utilised in the literature-Frijters et al. (2013) use handedness as an instrument of a child's ability, the validity of which has been contested (Grätz and Torche, 2016). Leight (2017) uses grain yields as a plausible instrument for differences in ability proxied by height-for-age. There is an extremely careful literature that has analysed whether parents compensate or reinforce specific (plausibly exogenous) policies and events experienced in childhood (Halla and Zweimüller, 2014; Adhvaryu and Nyshadham, 2016), which is highly informative, but may only be generalisable to larger policy shocks, whereas our use of variation in rainfall could be seen as 'normal' shocks to childhood experienced by children in low-income countries (Maccini and Yang, 2009).

Third, we descriptively examine heterogeneity in parental responses across socio-economic status, in a low-income setting. Such heterogeneity has been examined, but only in country contexts that are more developed than Ethiopia (Cabrera-Hernandez, 2016; Hsin, 2012; Grätz and Torche, 2016; Restrepo, 2016). To preview our results, we find that on average, parents provide more cognitive investment to the lower-ability child to reduce intra-household inequality. The compensatory parental responses appear to be concentrated in relatively higher-SES families. Specifically, we find suggestive evidence that families with educated mothers, smaller household size and higher wealth compensate through a higher level of cognitive investment when there are differences

[2] Other results on health in the study are based on a much larger sample.

in ability, while families with non-educated mothers, larger size and lower wealth exhibit only small and modest compensatory behaviours.

The paper proceeds as follows. In the next section we briefly review the relevant literature, and in subsequent sections then present our data, including the cognitive ability measures, followed by our econometric approach, our results and robustness checks and a concluding discussion.

2.2 Literature review

There are two competing theories on the direction of parental responses to observed ability in their children, both originating from theoretical models which are by now more than forty years old. Becker and Tomes (1976) predict that parents reinforce differences in child ability by investing more in the high-ability child, under the assumption that marginal return to investment is higher when the ability of the child is higher. In this case, parents' concern is for efficiency more than equity. On the contrary, Behrman et al. (1982) suggest that parents will compensate for ability differences to achieve equality among children when parents' inequality aversion preferences outweigh efficiency concerns.

In response, a burgeoning empirical literature has examined the effect of child endowments on parental responses. However, the results of this literature are mixed, indicating overall that there is either no clear direction of parental response on child endowment, or that the response depends heavily on context. Some studies have found evidence of reinforcing parental responses (Aizer and Cunha, 2012; Adhvaryu and Nyshadham, 2016; Behrman et al., 1994; Datar et al., 2010; Frijters et al., 2013; Grätz and Torche, 2016; Hsin, 2012; Rosales-Rueda, 2014); some have found compensating parental responses (Behrman et al., 1982; Bharadwaj et al., 2018; Cabrera-Hernandez, 2016; Del Bono et al., 2012; Frijters et al., 2009; Griliches, 1979; Halla and Zweimüller, 2014; Leight, 2017); some have found mixed responses (Ayalew, 2005; Hsin, 2012; Restrepo, 2016; Yi et al., 2015); some have found no effect at all (Abufhele et al., 2017; Almond and Currie, 2011).

Many of the recent empirical studies have relied heavily on the variation in birth weight to answer the question of parental responses, using a sibling fixed-effects (FE) model (Abufhele et al., 2017; Bharadwaj et al., 2018; Del Bono et al., 2012; Cabrera-Hernandez, 2016; Datar et al., 2010; Hsin, 2012; Restrepo, 2016; Rosales-Rueda, 2014). However, some studies argue that birth weight might be associated with prenatal endogenous input, and hence, exploit a source of exogenous variation in the endowment at birth. Halla and Zweimüller (2014) study the effect of an exogenous shock on the Austrian 1986 cohort, who experienced a prenatal exposure to radioactive fallout from the Chernobyl accident. The shock decreases the birth weight, live births and Apgar score; and increases premature births, and days for maternity leave. They find robust empirical evidence that parents compensate the children for experiencing input shocks. Adhvaryu and Nyshadham (2016) exploit variation in a plausible random *in utero* exposure to an iodine supplementation programme in Tanzania, and show that parents choose reinforcing investment in higher-ability children. Using Norwegian administrative data, Nicoletti et al. (2018) find that mothers compensate for low child birth weight by reducing maternal labour supply two years after birth. They instrument child birth weight by father's health endowment at birth, which arguably only brings variation in birth weight of child through genetic transmission without a direct impact on the mother's postnatal investments when conditioning on parental human capital and prenatal investments.

Meanwhile, other studies tackle this problem by using within-twins differences as a exogenous source of variation in endowment since prenatal parental investment is impossible to vary (Abufhele et al., 2017; Bharadwaj et al., 2018; Yi et al., 2015; Grätz and Torche, 2016). For example, Abufhele et al. (2017) find that parents are neutral to the difference in birth weight of twins in Chile and support the existing evidence that parents do not invest differentially between twins. Using the same data, Bharadwaj et al. (2018) find similar results that parents do not invest differentially within twins, while, using a sample of parents with singleton siblings, compensatory behaviour is found. As Almond and Mazumder (2013) noted, the reason could be that it might be especially costly for parents to implement differential treatment between twins.

Important concerns about using twins as an instrument have been raised. Using individual data in 72 countries, Bhalotra and Clarke (2018) find that the distribution of twins is not random in the population and that indicators of the mother's health and health-related behaviours and exposures are systematically positively associated with the probability of a twin birth. Certainly, twins are not a large proportion of the population, and may be seen more as a special case.

We build on two recent studies that examine the effect of child endowment on parental investment, and rather than relying on twins data, use instrumental variables to alleviate concerns of endogeneity bias resulting from both unobserved household heterogeneity and child-specific heterogeneity. Using sibling differences in handedness as an instrument for cognitive ability differences, Frijters et al. (2013) find reinforcing behaviours of parents who are more likely to allocate more cognitive resources on advantaged child in the USA. Grätz and Torche (2016), however, argue that handedness might vary over time so that it might not be an adequate instrument for child's early ability. Using the same technique but using variation in grain yields during the early life period of siblings as an instrument for physical health, Leight (2017) shows that Chinese parents invest more cognitive resources in the less-healthy child (as proxied by height-for-age) in Gansu province.

We combine a sibling-difference approach with instrumental variables, using the quasi-exogenous rainfall shocks occurring during the critical developmental period of a child as an instrument for differences in child ability between siblings.[3] As studies find that rainfall shocks have a substantial impact on child development in agricultural contexts (see Almond et al. (2018) for a review), we exploit differences between siblings by looking at rainfall shocks from *in utero* during the first three years of their life[4] as a source of exogenous variation in nutritional inputs during the critical development period experienced by the siblings.[5] Glewwe et al. (2001) note that a suitable instrument to capture within-sibling differences should be "(i) of sufficient magnitude and persistence to affect a child's stature; (ii) sufficiently variable across households; and (iii) sufficiently transitory not to affect the sibling's stature"(p.350). We provide robustness checks in this paper to argue that rainfall shock timing does indeed provide a plausible source of exogenous variation.

To our knowledge there are two other studies examining the pattern of parental investment in the context of Ethiopia. Ayalew (2005) examines catch-up growth of children in the dimensions of health and cognitive ability, using the first three rounds of the Ethiopia Rural Household Survey from 1994-95. He finds compensating behaviour in health, but reinforcing behaviour in cognitive skills. Arguably, the results for cognitive skills are less persuasive, since they use information on only one village in the survey.[6] Second, using the Young Lives Older Cohort data and relying on ordinary least squares (OLS)

[3]Rainfall information is external data matched with location by the Young Lives survey since the residence of interviewees is confidential.

[4]The period during pregnancy and the first 1000 days of life is widely recognised as the critical developmental period of child development (Doyle et al., 2009; Victora et al., 2010).

[5]Hill and Porter (2017) find that droughts cause a reduction in consumption of households in both rural and urban areas in Ethiopia.

[6]The outcome measure used is Ravens's Progressive Matrices scores, which did not work successfully in the Ethiopian context during Young Lives (Cueto and Leon, 2012) as children were unable to understand the task.

and fixed-effects (FE) estimations for identification, Dendir (2014) finds reinforcing behaviours, proxying parental investment with enrolment and child time allocation, and measuring ability using raw Peabody Picture Vocabulary Test (PPVT) scores[7]. Although the fixed-effects estimation successfully deals with the endogeneity issue caused by unobserved household characteristics, there is a potential high degree of correlation between child ability and unobserved child heterogeneity, such as parental preferences over one particular child, which is an individual effect. Dendir (2014) measured PPVT scores at adolescence (age 12 and 15), which increases the probability that this measure of ability is contaminated by unobserved child characteristics and consequently biases the results, and therefore exogenous variation in cognitive ability is necessary for more plausible estimation.

While most of the existing literature reveals how parents respond to the difference in health within siblings, to the best of our knowledge, only the two studies discussed above (Ayalew, 2005; Frijters et al., 2013) have examined differences in cognitive ability, and both have limitations. As it is of interest to show the specific parental response to one dimension of human capital, one would ideally like to disentangle the effect of investment in that particular dimension of human capital. However, constrained by data, only a few empirical studies have specific measures of investment in different dimensions, while most of the existing studies use a general measure of parental investment, such as time spent with the child. Yi et al. (2015)'s theory predicts that given the same early health shock, parents respond differently along different dimensions of human capital. The data they use contain detailed information on investment in family health and education. Yi et al. (2015) find mixed results: while parents compensate for the harmful effect of an early health shock by devoting more health resources to the worse-health child, they reinforce

[7]We discuss a better measure of ability and parental investment in Section 3.

in the domain of cognition by allocating fewer educational resources to the disadvantaged child. Restrepo (2016) and Rosales-Rueda (2014) use the same dataset from the USA, the National Longitudinal Survey of Youth-Children 1979 (NLSY-C79), which gives information on inputs of time and goods in either cognitive or socio-emotional development. They suggest that parents tend to simultaneously reinforce the effect of low birth weight by providing less cognitive stimulation and emotional support to the low-birth-weight child. In our study, we measure direct cognitive investment using total expenditure on educational fees at the individual child level.

Most existing research attempts to examine parental responses to child endowments on average. Some sociological studies emphasise that in theory, socio-economic heterogeneity should be taken into account, specifically, the degree and direction of parental responses might vary by family socio-economic status (SES) (Lareau, 2011; Lynch and Brooks, 2013). Some consider that lower-class parents have difficulty in affording costly and risky investment in disadvantaged children, and would be more likely to reinforce differences in ability. Higher-class parents tend to be averse to inequity so may compensate for a low ability outcome (Conley, 2008). On the contrary, others suggest that high-SES families may reinforce gaps in child ability by providing more educational investment to the advantaged child, while offering direct transfers, such as gifts or bequests, to the disadvantaged child (Becker and Tomes, 1976; Becker, 1991).

To date, only a small number of empirical studies have looked at variation in parental responses by SES, though these are all in a developed country context. Grätz and Torche (2016) find out that advantaged parents allocate more cognitive stimulation to higher-ability children, while disadvantaged parents behave indifferently to ability gaps. Yet, Halla and Zweimüller (2014) show that families with low socio-economic status chose to give birth to fewer children when their children experienced the Chernobyl accident; similarly, families with high socio-economic status compensate their low-endowed children by supplying less maternal labour (and investing more in childcare). Hsin (2012) uses maternal educational level to measure family socio-economic status. On average, no compensating or reinforcing investment is found for low birth weight outcomes. However, low-educated mothers prefer reinforcing investment by spending more time with heavier-birth-weight children at 6 years old, while high-educated mothers compensate low-birth-weight children by spending more time with them. Restrepo (2016) finds reinforcing behaviour on average, with low-SES families reinforcing the differences in birth weight with a greater amount of investment compared to high-SES families. None of these studies provide evidence in the context of developing countries, except Cabrera-Hernandez (2016) who finds that high-educated mothers in Mexico compensate for the low-birth-weight outcome by offering more school expenditure to the disadvantaged child.

2.3 Data and measures

Young Lives is an international longitudinal study of 12,000 children growing up in four developing countries (Ethiopia, India, Peru and Vietnam) over 15 years (Barnett et al., 2012), examining the causes and consequences of childhood poverty. The main cohort (2,000 children in each country) were born within 12 months of each other in 2001. An older cohort (1,000 children in each country) born seven years earlier is used as a comparison group. This paper uses data from four rounds of the Ethiopia survey, focusing on the Younger Cohort (YC) and their siblings. Round 1 was conducted in 2002 (when YC index children were, on average, 1 year old), Round 2 in 2006 (approximately age

5), Round 3 in 2009 (approximately age 8) and Round 4 in 2013 (approximately age 12). In Rounds 3 and Round 4, one sibling, closest in age to the YC index child (either younger or older), was interviewed. This brings variation in that YC index children could be either born earlier or later in our analysis.[8]

To reduce heterogeneity in child activities and parental investment, we confine the sample of YC index children[9] and their siblings to be aged from 7 to 14 in Round 4, being old enough to enter in primary school and young enough to stay in the primary school in Ethiopia. The sample is reduced to 701 sibling pairs (1,402 observations) in the sibling-difference specification, born from 1998 to 2006.

2.3.1 Rainfall

We use monthly rainfall data at community level additionally provided to us by Young Lives, which we merge with the survey data using birth year, birth month and birthplace (from Round 1 and Round 2 survey), in order to generate instrumental variables at the child-specific level. Annual rainfall is measured for each child from the 12 months prior to the birth month, so that the rainfall shock varies monthly and yearly. We use standardised annual rainfall from *in utero*, the first, second and third year of the child's life in the birthplace of the child as instrumental variable , following the literature arguing that this is the critical developmental period (Almond et al., 2018). During this period, adequate rainfall contributes to improved income for the household and therefore translates into a positive nutritional input for child ability (Maccini and Yang, 2009). The mean and standard deviation are calculated at the birth community level using rainfall from 1985 to 2008. In the context of Ethiopia, an extremely

[8]There are 610 YC index children older than their surveyed siblings, and 91 who are younger. The average age difference in month is 27 months. See Table 2.1 for details.

[9]In the following, we will describe YC index children as "index children" for simplicity.

drought-prone agricultural country, we hypothesise that higher the level of the rainfall during the critical developmental period, the better for the child's ability (Dercon and Porter, 2014).

FIGURE 2.1: Annual rainfall by community, 1998-2008

Since the sibling pairs in our sample are mainly born in the same community, the variation in the child-specific instrument variable relies on the time dimension, namely the birth month and birth year.[10] As the sibling pairs in the sample are born between 1998 and 2006, we check the distribution of annual rainfall in each community during the period from 1998 to 2008 (i.e. the second of year of life for a child born in 2006). Figure 2.1 shows that the rainfall in most of the communities is volatile, characterised by two severe droughts in 1999 and 2002 in Ethiopia. As 90% of the sibling pairs in our sample are born at least two years apart, the correlation of the rainfall instrumenting for

[10]Three percent of the sample were born in different communities due to migration, we include these and controls for community fixed effects. The results are robust to excluding this 3%, shown in Table 2.6 row 7.

each child ability is arguably weak.[11] Furthermore, we carry out a series of t tests to examine the difference in rainfall that sibling pairs experience in their early life respectively and find that the annual rainfall during the critical developmental period between index child and the sibling is statistically different. Specifically, the index child is reported to be exposed to a statistically lower level of rainfall as they are mostly born during 2001 and 2002, when drought hit Ethiopia.

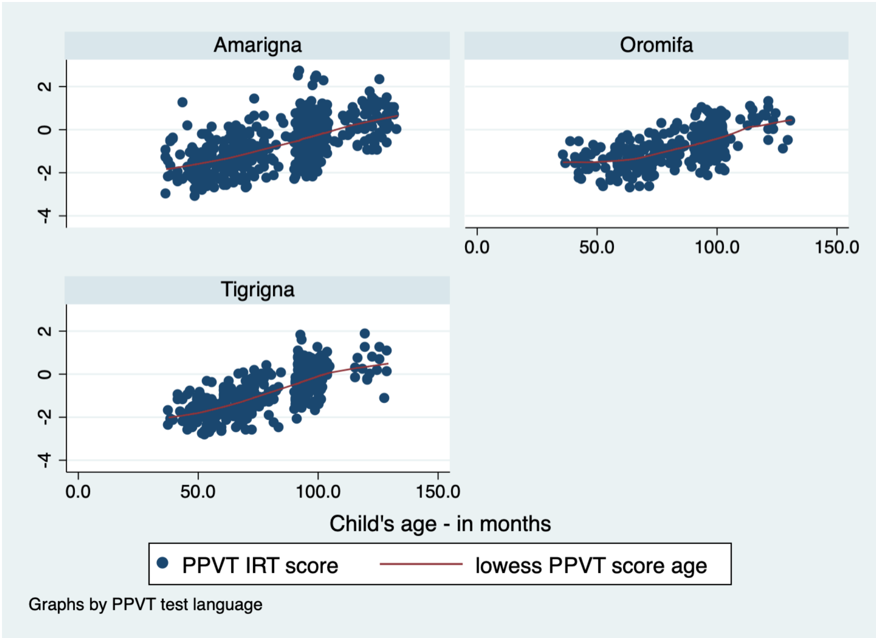
2.3.2 PPVT scores as a measure of cognitive ability

To analyse the effect of children's cognitive ability on within-household allocation of cognitive resources, our main independent variable of interest is the child's cognitive ability in 2009 (Round 3). The Peabody Picture Vocabulary Test (PPVT) is a receptive vocabulary test designed by Dunn and Dunn (1997), a consistent test measuring cognition ability for both index children and siblings in Young Lives. Therefore we measure the child's cognitive ability using this metric.[12] The PPVT is a widely used test to measure verbal ability and general cognitive development (see Crookston et al. (2013); Paxson and Schady (2007)), and the PPVT test score is positively correlated with other common measures of intelligence such as the Wechsler and McCarthy Scales (Campbell, 1998). Given that Round 3 is the first round that has information on siblings, our analysis only uses the latter two available rounds of the Young Lives data.

[11] We acknowledge that as a cohort study, half of the index children in our sample are born within twelve months of one another, and therefore if any policy or other shock which is correlated with rainfall happened during the birth period, then results may be influenced by such an unobserved cohort effect. However, examining the community datafiles from the Young Lives survey we did not find any such events.

[12] In the Young Lives study, there are two other cognitive tests, the Early Grade Reading Assessment (EGRA) and a maths test. However, they are only available for index children, not for siblings.

FIGURE 2.2: IRT PPVT scores by language, 2009



Given the difficulty of comparing raw PPVT scores across different rounds of data collection as children age, we employ item response theory (IRT) to standardise cognitive measures by language, following Leon and Singh (2017).[13] Figure 2.2 shows that the IRT PPVT scores increase along with the age, yet the means of IRT PPVT scores vary by language, consistent with findings of (Leon and Singh, 2017) (Tigrigna is the highest, followed by Amarigna and Oramifa). To ease the interpretation of subsequent estimation results, and given that is is not advisable to compare across languages (Cueto and Leon, 2012) the IRT scores have been standardised by language as our measure of cognitive ability, with a mean of 0 and a standard deviation of 1.

2.3.3 Total educational fees as a measure of cognitive resources

Our dependent variable is the allocation of parental cognitive resources, measured by the total education fees paid in 2013 (Round 4) per child. As

[13]See Leon and Singh (2017) for further details. We exploit the item parameters for each language calculated by (Leon and Singh, 2017) to generate IRT scores of children in Round 3. We use Stata command *openirt* programmed by Tristan Zajonc.

FIGURE 2.3: Young Lives Survey timings

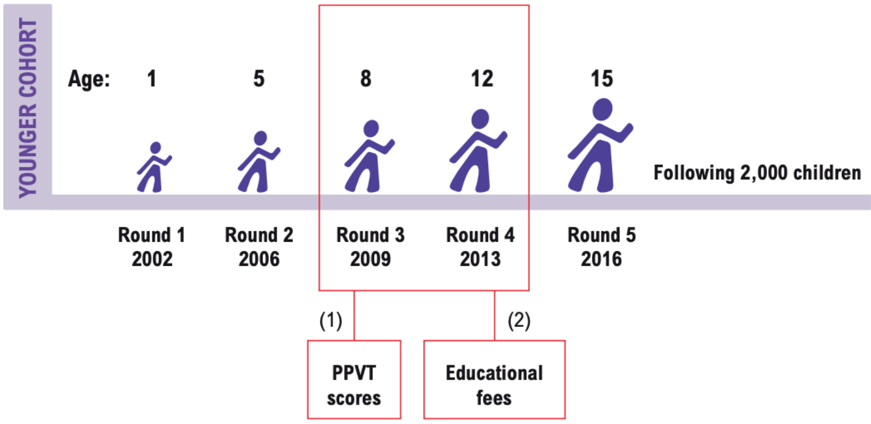


Figure 2.3 shows, an advantage of our panel data is that it leaves a longer period of time (four years between Round 3 and Round 4) to measure potential parental responses after children are assessed by PPVT in Round 3, while prior research mostly relies on the parental involvement measured quite soon after child ability is observed. The total educational fees are the sum of school fees and private tuition fees, serving as a direct measurement of cognitive investment.

To alleviate the concern that public educational investment and private tuition investment are substitute goods, we use Pearson's correlation[14] to test the strength and direction of the association between these two continuous variables. While the Pearson correlation coefficient between the school fees and tuition fees, $r = 0.732$ at 95% confidence level, suggests that in the pooled sample higher school fees are related to higher tuition fees, the correlation coefficient estimating the association between school fees and tuition fees within-family ($r = -0.020$) is statistically non-significant at 95% confidence level. This lack of correlation leads us to use total educational fees as the dependent variable of our main analysis.[15]

[14]We use Stata command *pwcorr* to carry out Pearson's correlation test.
[15] We provide analysis using private tuition fees as the dependent variable in the robustness check. See Table 2.6.

FIGURE 2.4: Total educational fees

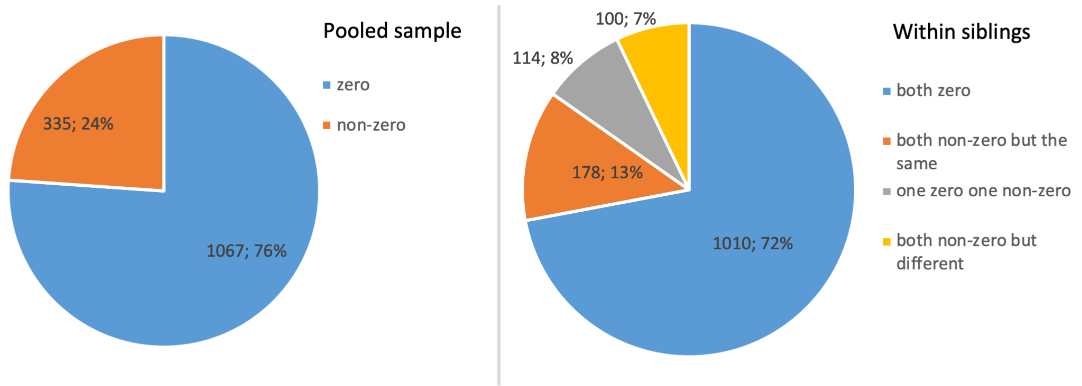


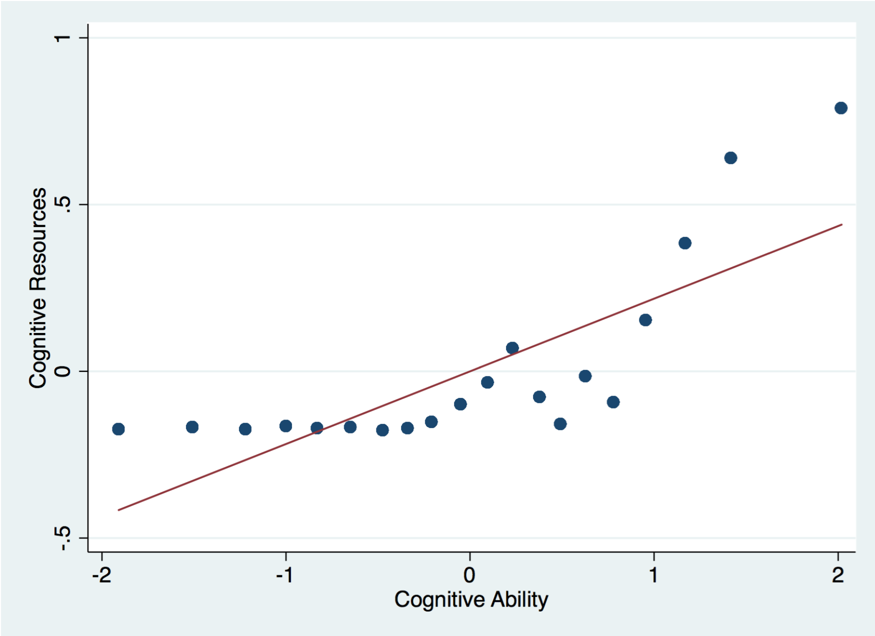
Figure 2.4 shows how total educational fees are reported. In the pooled sample, shown by the left-hand chart, 76% of parents report zero total educational fees in Ethiopia, while 24% report non-zero educational fees.[16] Looking at the allocation between siblings, indicated by the right-hand chart, 16% of parents differentiate their financial educational resources among their children, while 13% of parents allocate financial resources in child education and adopt no differentiating strategy in investing their children. Our interest is to find out whether the parental investing strategy of those who invest financial resources in their children is responsive to the difference in cognitive ability.

School fees and private tuition fees as a proxy of cognitive resources are specifically documented in parents' answers to the questions such as 'how much you spend on school (private tuition) fees per year?'. For the sake of interpretation, we standardise the total educational fees for the analysis.

To understand whether parents report a higher level of investment for the index children, we perform a t test on the total educational fees between index

[16]Our sample also includes those who are at school age but not enrolled currently, 121 children. We assign zero educational fees to them. The high percentage of zero educational fees is also due to the abolition of school fees in public schools for Grades 1 to 10 in Ethiopia in 1994. However, hidden costs remain (Oumer, 2009). UNICEF (2009) find that there were still payments in various forms in government schools after the policy of abolishing school fees. According to the Policy and Human Resource Development (PHRD) study, on average, a government school was levying about Birr 10 to 15 per year per student.

FIGURE 2.5: Mean cognitive resources and cognitive ability for each fifth percentile of the cognitive ability distribution



children and their siblings. The t statistics (= 0.132) shows that the difference in investment between two children is not statistically different, suggesting that parents do not deliberately report a higher investment for the index children.

Figure 2.5 shows the raw correlation between mean cognitive ability and mean cognitive resources for each 5 percentile for the included sample. Despite the flat relationship on the left tail of the distribution, the aggregate correlation between ability and parental investment is positive in the cross-section OLS estimation. Our interest is to find out whether this plausible positive relationship continues to hold when we apply our empirical methods accounting for child observable and unobservable factors.

Therefore, we include a series of child observable characteristics as confounding factors. First, to alleviate the concern that the cognitive investments are age-related, we control for several age-related factors in the regression analysis. We make use of age in months, together with square and cube of age in months and dummies of birth year. Then, since evidence suggests that children born earlier receive the greater investment (Price, 2008; Buckles and Kolka, 2014), we control for birth order. Other child-level differences which might contribute to investment variation are also controlled for in the regression analysis. Specifically, maternal age at birth, Height-for-age Z-score (HAZ) in Round 3, birthplace, birth quarter, and type of siblings (e.g., born as an older brother with a younger sister, or born as an older sister with a younger brother) are taken into account.[17] See Table 2.1 for summary statistics.

2.3.4 Socio-economic status (SES)

To understand whether educational investment varies by socio-economic status (SES), we carry out several exploratory *t* tests and find that families investing in education are indeed the high-SES families. The families who make positive investments in child education are significantly richer (*t*= -12.253), with a significantly better educated mother (*t*= -9.749) and smaller size (*t*=-3.991). In order to further investigate whether these better-off families who invest in education differentiate their investment based on the ability gap between their children, we stratify our analysis on parental responses. Specifically, we employ several household characteristics (maternal education, family wealth, and household size) as indicators of family SES, while we dichotomise each indicator generating a high-SES group and a low-SES group following Grätz and Torche (2016). With regard to maternal education, in fact, half of the mothers in our sample are not educated at all, so that we distinguish between families by having an educated mother or a non-educated mother. In the case

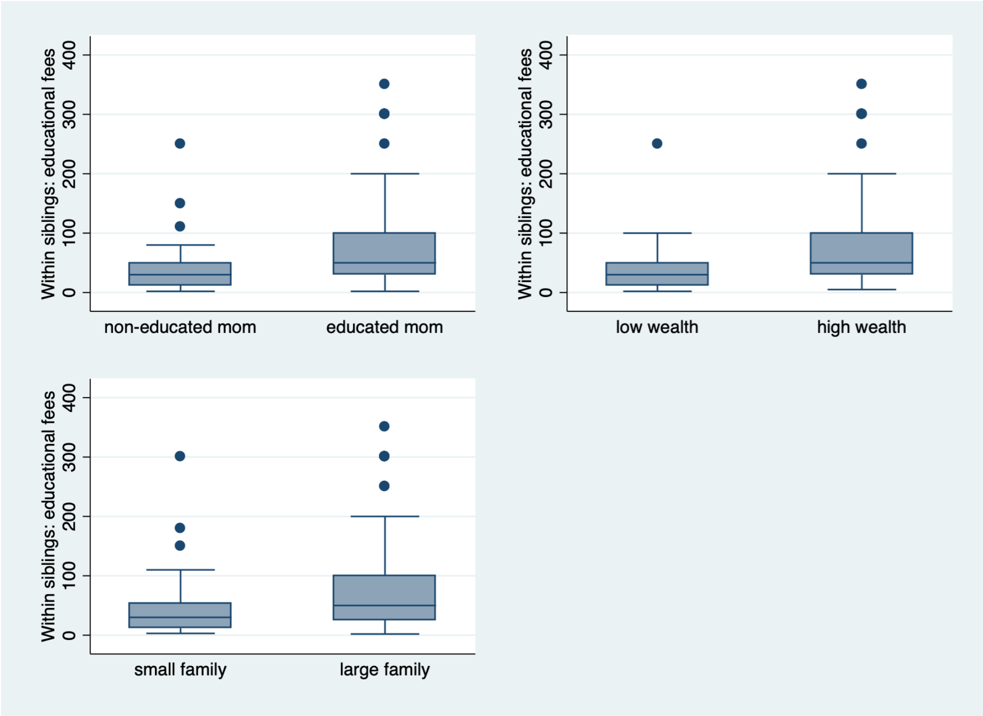
[17]There are eight factor variables to denote the type of siblings: born as an older brother with a younger sister, born as younger sister with a older brother, born as an older sister with a younger brother, born as a younger brother with a older sister, born as an older brother with a younger brother, born as younger brother with a older brother, born as an older sister with a younger sister, and born as a younger sister with a older sister. When we use our fixed-effects strategy, many are dropped due to their multicollinear relationship when the information of index children is deducted by their siblings'. Note that we only include time-varying household characteristics due to our sibling difference specification

TABLE 2.1: Summary statistics

Variable	Mean	SD	Mean (within)	SD (within)
Cognitive resources				
Total educational fees (standardised)	0.000	1.000	-0.007	0.054
Cognitive ability				
PPVT scores (standardised)	0.000	1.000	-0.867	0.891
Child characteristics				
Age in months	131.758	21.263	-26.765	23.184
Maternal age in months at birth	27.370	6.064	2.215	1.952
Birth order	3.490	1.858	0.743	0.682
Born as an older sister with a younger brother (dv)	0.118	0.322	-0.173	0.454
Born as an older brother with a younger sister (dv)	0.135	0.342	-0.218	0.471
Born as an older brother with a younger brother (dv)	0.130	0.336	-0.211	0.464
Born as an older sister with a younger sister (dv)	0.118	0.322	-0.138	0.465
Born as a younger brother with an older sister (dv)	0.130	0.336	0.211	0.464
Born as a younger sister with an older brother (dv)	0.118	0.322	0.138	0.465
Born as a younger brother with an older brother (dv)	0.118	0.322	0.173	0.454
Born as a younger sister with an older sister (dv)	0.135	0.342	0.218	0.471
HAZ in Round 3	-1.359	1.129	-0.141	1.233
Birth quarter 1 (dv)	0.213	0.410	0.013	0.577
Birth quarter 2 (dv)	0.295	0.456	-0.054	0.646
Birth quarter 3 (dv)	0.216	0.412	0.021	0.608
Birth quarter 4 (dv)	0.275	0.447	0.020	0.618
Rainfall *in utero* (standardised)	0.075	0.901	-0.497	1.210
Rainfall at birth (standardised)	-0.449	1.039	1.225	1.440
Rainfall in year 1 (standardised)	-0.155	0.796	0.924	0.997
Rainfall in year 2 (standardised)	-0.055	0.704	0.307	0.956
Household characteristics				
Wealth index	0.348	0.164	0.000	0.000
Mother with education (dv)	0.459	0.499	0.000	0.000
Household size	6.522	1.646	0.000	0.000
N	1402			

Note: 'dv' is denoted for dummy variables; 'within' stands for the data constructed in 'within-family' structure.

FIGURE 2.6: Intra-household difference in total educational fees by SES



of family wealth and household size, we dichotomise them using the median of wealth index and size of the family.[18]

Figure 2.6 shows the intra-household difference in total educational fees by SES. The distributions of within-sibling difference in educational fees are similar across three indicators. In general, the high-SES families have bigger differences in allocating educational resources among their offspring. The mean of the difference in total educational fees in low-SES families is small but non-zero.

2.4 Econometric strategy

To identify the causal effect of cognitive ability on parental investment, the analysis is based on an IV-FE model, targeting three main endogeneity threats.

[18]The wealth index is the average of housing quality index, consumer durable index and housing service quality index.

First, this approach relates within-sibling pair differences in ability in 2009 (Round 3) with within-sibling pair differences in parental cognitive investment four years later in 2013 (Round 4) to address the threat of reverse causality. Second, the sibling fixed-effects (FE) models control for unobserved heterogeneity at the household level, following most existing empirical work. Third, we use instrumental variables to isolate the exogenous variation in child ability, addressing endogeneity resulting from unobserved child heterogeneity.

The sibling fixed-effects structural model can be written as follows:

$$\Delta I_h = \beta \Delta CA_h + \Delta X_h \Lambda + \Delta \epsilon_h \tag{2.1}$$

where ΔI_h is the difference in cognitive investment between siblings in family h in Round 4 (i.e., total educational fees), ΔCA_h is the difference in ability between siblings in Round 3, ΔX_h is a vector of differences in other characteristics between siblings (e.g. child's age, maternal age at birth, height-for-age in Round 3, birthplace, birth quarter, birth year, birth order, type of sibling pairs - gender of older and younger child), and $\Delta \epsilon_h$ is the difference of the idiosyncratic error term between siblings. In this estimation, time-invariant household observable characteristics and household unobservable confounding factors are purged from the specification, but unobserved child heterogeneity, such as personality, remains.

As noted above we overcome endogeneity bias resulting from unobserved child heterogeneity, with an instrumental variables (IV) estimation procedure. The first stage equation is:

$$\Delta CA_h = \sigma \Delta R_h + \Delta X_h \Omega + \Delta \mu_h \tag{2.2}$$

where ΔR_h is the difference in rainfall shock from *in utero* to the first three years of child's life between siblings as a source of exogenous variation in nutritional inputs experienced by the siblings, and $\Delta\mu_h$ is a random error term in the first stage.

IV approach is also helpful in the sense of overcoming attenuation bias related to measurement error in cognitive ability. Even if we consider the PPVT test score a good proxy for ability observed by parents, there is still likely to be measurement error in the test, and in its relation to parental perception of ability. For example, parents may have some other perception of their children's cognitive ability than the PPVT score. This potential problem of measurement error can be solved by our IV approach if it is classical. Indeed, in a sibling FE model, attenuation bias caused by measurement error is augmented if one's analysis moves from a cross-sectional setting to a FE setting (Bound and Solon, 1999).

The sibling FE model coupled with the IV strategy helps us to interpret β as the Local Average Treatment Effect of change in parental cognitive investment caused by the variation in child cognitive ability, which is driven by the exogenous variation in rainfall during the critical developmental period of the two children. We note that the monotonicity assumption applies to LATE estimates (Angrist and Pischke, 2008), that for an change in rainfall, there should be a monotonic increase in "treatment" intensity. If $\beta > 0$, parental investment increases with relative ability. Parents reinforce the differences in ability by allocating more resources to the high-ability child. If $\beta < 0$, it means parents compensate for the difference in ability, allocating more resources to the low-ability child.

Under the assumption of higher marginal returns to investment in higher-ability children, the case of $\beta > 0$ also implies that parents are concerned more with the efficiency of investment and try to maximise their children's total future wealth. The case of $\beta < 0$, on the other hand, implies that when equity outweighs efficiency, parents forgo maximising returns from educational investment, trying to achieve higher equity among children. Del Bono et al. (2012) note also that there may also be a "pure endowment effect", whereby if marginal utility of parents with respect to any individual child's ability is positive but decreasing (i.e., the second derivative of the utility function is negative), then higher endowment of that child effectively increases family resources, but these can only be released by decreasing investment in that child. This effect is then expected to operate in the same direction as the equity effect.

We report two types of standard errors, one robust to general heteroskedasticity and the other one robust to within community dependence.[19]

2.5 Results

In order to test the relationship between cognitive ability and deployed cognitive resources, we regress parental cognitive resource allocation in primary school on cognitive ability observed one period earlier. In all of the estimation results, total educational fees paid for each child is the proxy for cognitive resources, while PPVT scores are the proxy for cognitive ability.

For each specification we use the sample of children who have a surveyed sibling and the information for both siblings is available. Furthermore, we have restricted the sibling-pairs to be at primary school age and use the same language in PPVT test. A set of child-level covariates are included in all models, such as age in months, maternal age at birth, height-for-age in Round 3, birthplace, birth quarter, birth order, type of sibling pairs and birth year.

[19]There are 46 clusters in the sample.

2.5.1 Preliminary results

Table 2.2 presents the preliminary results from the OLS models and FE model. The inconsistency of the estimates from these models is evident, the magnitudes and signs of which are not stable as we add additional controls, suggesting severe endogeneity of the variable of interest. For example, the cross-sectional OLS estimate reported in column 1, when only child-level controls are included in the model, suggests a positive relationship between ability and total educational fees. However, when we include household-level traits, maternal educational background and regional fixed-effects in the model, the point estimate decreases from 0.133 to 0.059.

However, the OLS estimate is still likely to be biased due to unobserved characteristics within the family, such as genetically innate ability, parental preferences for child quality, and budget constraints. Hence, we exploit the sibling-FE model, using a similar strategy to Bharadwaj et al. (2018), Datar et al. (2010) and Hsin (2012) studying parental responses to birth-weight, controlling for unobserved household-level characteristics. In column 5 of Table 2.2, the FE estimate suggests a negative association between ability and investment, although it is not statistically significant. Aside from this, endogeneity bias might still persist since the cognitive ability is postnatal and time-varying, which allows after-birth ability to embody a significant component of prior parental investment.

To address the bias, we isolate the exogenous variation in cognitive ability using quasi-exogenous variation in rainfall during the critical developmental period. Thus we apply instrumental variable methods to the sibling fixed-effects approach (IV-FE), a similar approach to Frijters et al. (2013) and Leight (2017), who use the same strategy but different instruments to ours.

TABLE 2.2: Preliminary regression models

DEPENDENT VARIABLE:	OLS	OLS	OLS	OLS	FE
TOTAL EDUCATIONAL FEES	(1)	(2)	(3)	(4)	(5)
Cognitive Ability	0.133**	0.076*	0.064**	0.059**	-0.002
	(0.060)	(0.041)	(0.032)	(0.029)	(0.004)
Child-level controls	Yes	Yes	Yes	Yes	Yes
Household-level controls	-	Yes	Yes	Yes	Yes
Mother-level controls	-	-	Yes	Yes	Yes
Region fixed-effects	-	-	-	Yes	Yes
Sibling fixed-effects	-	-	-	-	Yes
Observations	1402	1402	1402	1402	1402

Note: Community clustered standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is total educational fees. Children controls are age in months, square of age in months, cubic of age in months, maternal age at birth, gender, birth place, birth quarter, birth order, birth year, height-for-age Z-score, language of tests, and the type of sibling. Household-level controls are type of residential site, household size, wealth index, and gender of household head. Mother-level controls are a series of levels of maternal education.

2.5.2 Main results

2.5.2.1 IV-FE models: First-stage results and diagnostics

Before presenting our main IV-FE results, we discuss the first-stage results, as well as the underidentification and weak identification tests in Table 2.3. Specifically, in the first-stage estimations endogenous cognitive ability is regressed on the exogenous regressors and excluded instruments (i.e., the rainfall during critical developmental period). We find that children who experienced relatively good rainfall aged 0-24 months have significantly higher test scores than their siblings in their early childhood; rainfall during infancy is *relevant* to cognitive ability as proxied by receptive vocabulary.

Shown in column 1 to 4 in Table 2.3, we regress ability in childhood on annual rainfall from *in utero* to the first three years of child life respectively. We find that annual rainfall during 0 to 12 months of life and 13 to 24 months of life are significant. Therefore, we construct an IV using the average rainfall during 0 to 24 months of life and report the result in column 5. The estimate is positive and statistically significant, with a t statistic of 5.70, suggesting that an increase of one standard deviation in rainfall during the first two years of life is correlated with an increase of 15.6% of one standard deviation in cognitive ability in early childhood. In column 6, when we include both the rainfall during the first year and the second year of life as IVs into the IV-FE model, both of the estimates are positive and statistically significant.

With regards to the underidentification tests[20], the p-values for the specifications 2, 3, 5, and 6 all reject the hypothesis that the IV models are underidentified respectively, though not specifications 1 and 4, suggesting that the IV models are likely to be underidentified using either rainfall *in utero* (column

[20]The underidentification test is an LM version of the Kleibergen and Paap (2006), which allows for non-i.i.d. errors.

TABLE 2.3: First stage regressions: Results and tests of underidentification and weak identification

Cognitive Ability	(1)	(2)	(3)	(4)	(5)	(6)
Rainfall *in Utero*	-0.038					
	(0.027)					
	[0.028]					
Rainfall at Birth		0.102				0.069
		(0.020)***				(0.023)***
		[0.028]***				[0.029]**
Rainfall in Year 1			0.144			0.091
			(0.029)***			(0.033)***
			[0.042]***			[0.044]**
Rainfall in Year 2				0.047		
				(0.034)		
				[0.034]		
Average Rain at Birth and Year 1					0.156	
					(0.027)***	
					[0.038]***	
Underidentification test:	$\chi(1)^2 = 2.092$	$\chi(1)^2 = 24.543$	$\chi(1)^2 = 24.472$	$\chi(1)^2 = 2.011$	$\chi(1)^2 = 30.805$	$\chi(2)^2 = 31.132$
p value	0.148	0.000	0.000	0.156	0.000	0.000
Weak instrument test:						
Montiel-Pflueger (MP) effective F stat	2.012	25.582	24.716	1.958	32.539	17.169
Montiel-Pflueger critical values:						
5% of worst case bias	37.418	37.418	37.418	37.418	37.418	5.808
10% of worst case bias	23.109	23.109	23.109	23.109	23.109	4.550
20% of worst case bias	15.062	15.062	15.062	15.062	15.062	3.828
Observations	1402	1402	1402	1402	1402	1402

Note: Within-household fixed effects estimates. Robust standard errors in parentheses. Clustered standard errors by community in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Child controls include age in months, square of age in months, cubic of age in months, maternal age at birth, birth order, height-for-age Z-score in round 3, birthplace, birth quarter, birth year, and the type of sibling (such as born as an older sister and paired with a younger brother). Both the underidentification test and weak instrument test are robust to heteroskedasticity. The Montiel-Pflueger (MP) F statistics are very similar to Kleibergen-Paap rk Wald F statistics in weak instrument test. The MP weak instrument test offers valid critical values at 95% confidence level and test statistics in the absence of assumption of i.i.d. data.

1) or rainfall in the third year of life (column 4) as the excluded instrument variable. Therefore, in the following, we focus on four IV-FE models: three of them are single IV models using rainfall from 0 to 12 months, rainfall from 13 to 24 months, and average rainfall from 0 to 24 month; the last one is a two-IV model using both rainfall from 0 to 12 months and from 13 to 24 months.

We further examine the validity of the IVs by conducting a battery of weak identification tests. Noting that the traditional Cragg-Donald weak instrument test applies to the case of i.i.d. data only, we report a robust weak instrument test by Olea and Pflueger (2013) which gives valid test statistics -Montiel-Plueger (M-P) effective F statistics- and Montiel-Plueger critical values in the existence of heteroskedasticity at 95% confidence level.[21] Although the robust M-P F statistics in the specifications 2 and 3, which are 25.582 and 24.716, satisfy the "rule of thumb" recommended by Staiger and Stock (1997), when comparing them with the robust critical values given by M-P test, we notice that these statistics are slightly higher than the M-P critical value for a maximum IV bias of 10%, suggesting that there is a 5% chance that the bias in the IV estimator is 10% of the worst case possible.

When we use average rainfall between the age of 0 to 24 months as the excluded instrument, the robust weak instrument test suggests that this IV is reasonably "strong". In column 5 in Table 2.3, the robust M-P F statistic is 32.539, which is sufficiently above the robust M-P critical value for a maximum IV bias of 10%. Additionally, the combined set of instruments in column 6 are also "stronger" than the ones in column 2 and 3, as its M-P F statistic is 17.169 which is higher than the critical value for a maximum IV bias of 5%.

[21]We use *weakivtest* programmed in Stata by Pflueger and Wang (2015). The Montiel-Plueger effective F statistics are very close to the built-in Kleibergen-Paap rk Wald F statistics in the programme *xtivreg*2 wrtten by Schaffer (2015). However, the robust critical values of the latter are not provided. Thus we use the Montiel-Plueger critical values as thresholds in order to report the bias.

To conclude, the single instrument of average rainfall at the age from 0 to 24 months and combined set of instruments of rainfall at the age from 0 to 24 months are respectively *relevant*, implying the second-stage inferences will be valid and point estimates are only likely to include a relative bias lower than 10% at 95% confidence level. These results could also serve as a supplement to the studies investigating whether some periods during the critical developmental period are more important. We find that children are particularly vulnerable at the age of 0 to 24 months in developing cognitive ability in Ethiopia, which is consistent with the findings of Maccini and Yang (2009), though Dercon and Porter (2014) find children exposed to famine at the age of 12 to 36 months are shorter than their peers in Ethiopia. In Table A3, we show IV redundancy test of a specified IV, which supports the hypothesis that rainfall at the age from 0 to 24 months are not redundant.

2.5.2.2 IV-FE models: Second-stage results

Our main estimation results are presented in Table 2.4, where the second-stage estimations using four IV models selected from above are presented. Across the four IV-FE models, the point estimates[22] are significantly negative, suggesting a compensating behaviour when parents observe their child to be under developed. Particularly, remembering that the preferred IVs used in specification 5 and 6 in Table 2.3, whose corresponding results are shown in column 3 and 4 in Table 2.4, are relatively more *relevant*, the point estimates of these two specifications are very close (-0.038 and -0.039). It suggests that an increase in cognitive ability of one standard deviation decreases cognitive resources by 3.8%-3.9% of a standard deviation.[23]

[22]The IV-FE point estimates are given by *xtivreg2* programmed by Schaffer (2015).
[23] The mean of the total educational fees is 123.23 Birr.

TABLE 2.4: IV-FE regression models of cognitive ability and total educational fees

DEPENDENT VARIABLE: TOTAL EDUCATIONAL FEES	IV-FE			
	Instruments: Rainfall at birth	Instruments: Rainfall in year 1	Instruments: Average rainfall in the first two years of life	Instruments: Rainfall at birth & rainfall in year 1
	(1)	(2)	(3)	(4)
Cognitive ability	-0.034	-0.045	-0.038	-0.039
	(0.015)**	(0.014)***	(0.013)***	(0.013)***
	[0.017]**	[0.019]**	[0.017]**	[0.017]**
Anderson-Rubin (AR) test	[-0.064,-0.011]	[-0.075,-0.025]	[-0.064, -0.019]	[-0.075, -0.019]
p value	0.016	0.000	0.001	0.001
Moreira CLR test	-	-	-	[-0.067,-0.023]
p value				0.000
K test	-	-	-	[-0.066,-0.023]
p value				0.000
J test	-	-	-	entire grid
p value				0.411
K-J test	-	-	-	[-0.069,-0.022]
p value				0.000
Observations	1402	1402	1402	1402
No. Excluded Instruments	1	1	1	2

Note: Robust standard errors in parentheses, community clustered standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is standardised total educational fees. Child controls are age in months, square of age in months, cubic of age in months, maternal age at birth, birth order, height-for-age Z-score in round 3, birthplace, birth quarter, birth year, and the type of sibling. The AR test, CLR test, K test, J test and K-J test are all robust to heteroskedasticity. All the tests give confidence intervals at 90% confidence level. The AR test and K-J test are a joint test of the structural parameter β and the exogeneity of the instruments, where K and CLR only test the former. The K-J test is more efficient than the AR test. K test and CLR test are more powerful than AR test. Unlike the K test, the K-J test and CLR test do not suffer from the problem of spurious power losses. The J test is like the Hansen J test of weak exogeneity, giving a confidence set where all values of β that are consistent with the assumption of weak exogeneity of instrument variables.

The confidence intervals given by a set of weak identification tests [24] are negative. Specifically, the Anderson-Rubin test (AR) gives negative confidence sets of estimated β that is robust to potential bias introduced by weak instruments gives negative confidence intervals at 90% confidence level. In the two-IV model, besides the AR test, the Moreira CLR test, K test, and K-J test are available. The K-J test is more efficient than the AR test, and Moreira CLR test and K test obtain more power than AR test when the model is over-identified. Compared to K test, K-J test and Moreira CLR test do not suffer from spurious power losses (Finlay et al., 2016). While all of the AR, Moreira CLR, K, and K-J tests give negative confidence sets, the latter three are almost identical, between [-0.067, -0.023] at 90% confidence level. The J test rejection probability is low everywhere except for very high values of β, suggesting that the instruments are exogenous.

To allay the concern of our proposed IV being possibly not perfectly *exogenous*, we further exploit a newly developed estimator by Conley et al. (2012), which identifies a threshold for the plausible estimate even if the IV is *imperfect*, i.e., the excluded instrument is directly correlated with the dependent variable.[25] Specifically, one might worry that rainfall in infancy might have a direct impact on contemporaneous parental investment, despite our argument that the impact on household income is only contemporaneous and short-lived (Glewwe et al., 2001); another concern might be that the early life rainfall would affect early life parental responses, which are auto-correlated with contemporary parental responses–our outcome variable. We argue that if such auto-correlation exists, the direction will be positive, if parental investment strategy is consistent over time, in particular we assume that parents would

[24]The AR, Moreira CLR, K, J and K-J confidence intervals are given by *weakiv*, programmed by Finlay et al. (2016).

[25]We use *plausexog* programmed in Stata by Clarke (2017), using the union of confidence interval approach for estimation of bounds.

not switch from reinforcement in early life to compensation in later life. Therefore, if early compensation effort exists, the difference of parental investment would be negatively correlated with the difference of rainfall in early life. To generate a robust estimate under this prior (Conley et al., 2012), we allow departures from the assumption of strict exogeneity of rainfall so that rainfall could have a non-zero and direct impact on parental investment, whose size is in the interval of $[-\delta, \delta]$.[26] As shown in Figure B1, we identify the lower bound of the direct effect which would render the second-stage estimate of the interest parameter insignificant at 10% confidence level. The results show that if the lower bound is greater than -0.003, the second-stage estimate would be significant. As the overall reduced-form effect of rainfall on parental response is -0.0059, we are confident that the lower bound of the model still is significant, given that the direct effect would have to be greater than 51% of the overall effect to render the IV point estimate insignificant.[27]

To further allay the concern of rainfall having a long-term effect on consumption, in the robustness check shown in Row 11 of Table 2.6, we also provide results on whether the idiosyncratic rainfall during the period when children are aged between 0 and 2 years old has a direct impact on consumption in the future at household level, and this is not significant.

Another related threat to the exclusion restriction would arise if rainfall in one siblings infancy affects the other's ability (earlier or later than the critical periods in question). Therefore, we regress ability on rainfall exposure of both own and sibling rainfall shock in infancy, while replacing the household fixed

[26]Following Conley et al. (2012)'s "plausible exogeneity" test, we propose a model derived from equation (2.1), $\Delta I_h = \beta \Delta CA_h + \gamma \Delta R_h + \Delta X_h \Lambda + \Delta \epsilon_h$, where difference in rainfall has a non-zero impact on parental responses. In the conventional IV approach, γ is set to be zero. Conley et al. (2012) note that in theory if we know γ we could subtract it from both sides of the equation and continue with a consistent IV estimate. They relax this restriction so that γ is not necessarily zero, but in the bounds of $[-\delta, \delta]$, allowing us to see whether this direct effect is large enough to render the IV estimate insignificant.

[27]The overall reduced-form is estimated using the model $\Delta I_h = \gamma \Delta R_h + \Delta X_h \Lambda + \Delta \epsilon_h$.

effects by county fixed effects since estimating coefficients on own and sibling rainfall exposure would not be possible in a family fixed effects model. Shown in Table A2, the estimates of rainfall during infancy of the sibling are insignificant and the magnitude is as small as a tenth of the one of our interest variable (in absolute value). The coefficient of child's "own" rainfall during the first two years of life remains significant and large in magnitude after including the sibling rainfall.

Finally, we consider that parents' education decision may be influenced not only by cognitive development but also by the child's health. To address this concern we add current health to the vector of controls. Removing HAZ does not change our results, as shown in Row 10 of Table 2.6. We also considered health as an alternative main variable of interest, given that early rainfall may also affect health, and nutrition can be proxied by height-for-age. We therefore reran our main model with HAZ in 2009 as the proxy for child "ability". Rainfall was a weak instrument for HAZ, and the second stage results were insignificant.[28] This echoes the health literature which shows that children may recover from early height deficits by mid-childhood, but cognitive ability in mid-childhood is still highly correlated with early nutrition (Casale and Desmond, 2016).

We compare our results with others using the IV-FE approach to examine parental responses. We noted some limitations of Frijters et al. (2013) handedness instrument earlier. In addition, the traditional Cragg-Donald F statistic of 12.32 under the assumption of an i.i.d. error, only satisfies the "rule of thumb" marginally, and arguably fails to provide strong evidence that handedness is a valid instrument to identify child's ability. On the contrary, Leight (2017)'s grain yield instrument is robust to the existence of weak instrument using p value from an AR test.

[28]See Table A5 for full results.

More generally, our finding of a strong negative relationship between cognitive ability and cognitive resources is consistent with a number of studies finding that parents prefer inequality aversion (Behrman, 1988; Bharadwaj et al., 2018; Rosenzweig and Wolpin, 1988; Del Bono et al., 2012; Frijters et al., 2009; Halla and Zweimüller, 2014; Leight, 2017; Yi et al., 2015).

2.5.3 Heterogeneity of parental responses to children's early ability

After studying the parental response at an aggregate level, we now explore heterogeneities in responses by stratifying the sample by maternal education, household size and wealth. Splitting the sample according to endogeneous characteristics is not an ideal solution, however, we follow the existing literature on this topic for more developed countries than Ethiopia (Cabrera-Hernandez, 2016; Hsin, 2012; Grätz and Torche, 2016; Restrepo, 2016), given that the heterogeneous characteristics we are interested in are fixed at the household level, so these cannot be interacted in the IV model; we interpret our results here with some caution.

Table 2.5 suggests that the association between early ability and later cognitive resources does vary by family socio-economic standing (Models 2-7). Specifically, the point estimates show high-SES parents provide more cognitive stimulation to their low-ability child, whereas low-SES parents compensate less cognitive investment in ability between their children. This heterogeneous variation in parental responses across SES is consistent using three indicators of socio-economic standing (maternal education, household size and family wealth). We should emphasise that the Young Lives sample are already a "pro-poor" sample from communities that are relatively poor, in a country that is poor by global standards (Outes-Leon and Sanchez, 2008).

TABLE 2.5: IV-FE model of the effects of cognitive ability on total educational fees: potential heterogeneity effect by SES

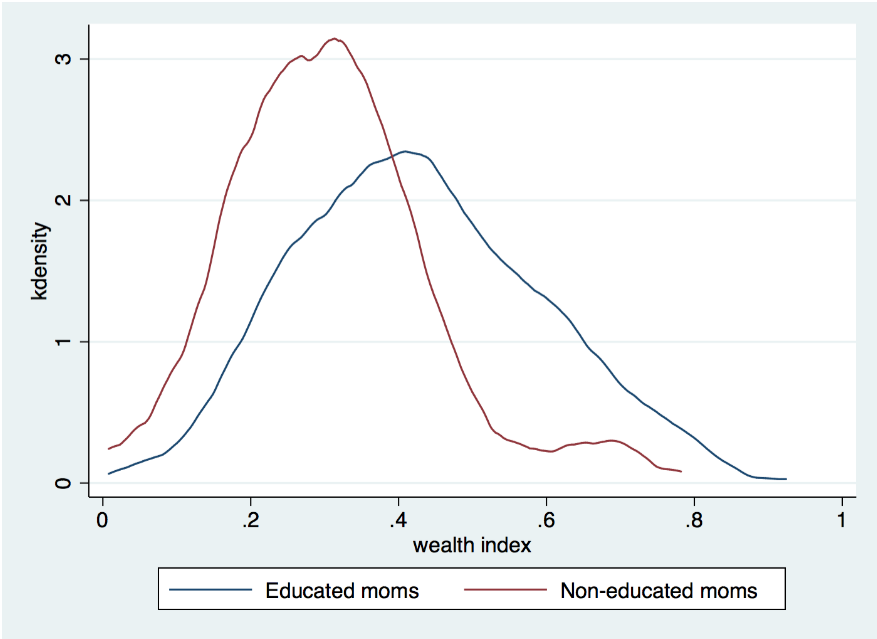
		Maternal Education		Household Size		Family Wealth	
	All	No	Yes	Large	Small	Low	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cognitive Ability	-0.039	-0.024	-0.053	-0.027	-0.053	-0.021	-0.063
	(0.013)***	(0.009)***	(0.025)**	(0.012)**	(0.023)**	(0.008)***	(0.028)**
	[0.017]**	[0.013]*	[0.028]*	[0.013]**	[0.032]	[0.012]*	[0.033]*
Weak instrument test:							
MP effective F stat	17.169	8.961	8.984	8.062	9.531	7.268	10.209
MP critical values:							
5% of worst case bias	5.808	7.153	11.130	8.111	7.805	7.502	12.098
10% of worst case bias	4.550	5.333	7.678	5.892	5.721	5.542	8.247
20% of worst case bias	3.828	4.276	5.668	4.601	4.505	4.400	6.010
Wald test of estimates:							
$\chi(1)^2$		1.18		0.98		2.01	
Prob > χ^2		0.272		0.321		0.157	
Observations	1402	758	644	664	738	792	610

Note: Robust standard errors in parentheses. Clustered standard errors by community in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.Same model is used as the main model. The IVs used are rainfall at birth and rainfall in year 1.

Table 2.5 shows that among better-off parents (educated mothers, small household or high family wealth), a one standard deviation increase in ability leads to 5.3% to 6.3% of a standard deviation decrease in total educational fees, while worse-off parents only compensate 2.1% to 2.7% of a standard deviation more educational investment to the low-ability child. Given the size of the standard errors, there is some overlap in the 95% confidence intervals for the point estimates, which means we cannot conclude definitively that they are significantly different, but given the fairly small sample sizes, we do consider the evidence as strongly suggestive. In comparison, using a sibling FE model, Hsin (2012) and Restrepo (2016) suggest a compensating effect among high-educated mothers by providing more time and more cognitive and emotional stimulations to the low-birth-weight children in the USA. The result is also consistent with Cabrera-Hernandez (2016) which uses a sibling FE model and finds out higher-educated mothers compensate expenditure in school for the low-birth-weight outcome in Mexico. However, Grätz and Torche (2016) use a twin FE model and find that advantaged families provide more cognitive stimulation to higher-ability children, and lower-class parents do not respond to ability differences in the USA.

An unanswered question based on the existing findings is that whether the heterogeneous result by maternal education is caused by the difference in wealth, in differential preferences for compensation, or ability to observe a difference in the cognitive outcomes of the siblings. Figure 2.7 shows that on average, mothers with education are generally better off in terms of wealth, implying that educated mothers might have a higher capacity to compensate disadvantaged children, simply because they have sufficient financial resources.

FIGURE 2.7: Kernel density plot of household wealth index by maternal education, 2013



2.5.4 Robustness checks

We now present some additional robustness tests. First, we restrict the sibling-pairs to have an age gap larger than two years, i.e., the older sibling should be born at least three years earlier than the younger one, in order to avoid a direct relationship between the rainfall shock experienced by one and outcome of the other. For example, one could argue that if one child is born one year after the older sibling, the rainfall experienced by the older one in the second year of life would be the rainfall the next child experiences in the first year of life; also, when the new born child is exposed to an adverse shock at birth, the parent might reallocate the resources immediately among the children which would directly influence the nutritional input of the older child in the second year of life. The restricted sample has 844 observations. In Row 1 of Table 2.6, the first-stage coefficient of rainfall in the first two years of life equals 0.187 (t=5.67), which is only slightly larger than the one of full sample presented in Table 2.3. The full diagnostics of the first stage using restricted sample is shown by Table A4, which is consistent with the results of the full sample. The second-stage estimate equals -0.036 (z=-2.57), very close to the one in full sample which equals -0.038. To conclude, it is consistent with previous result that parents compensate for the disadvantaged children by offering higher educational resources to them. This can also serve as a suggestive evidence that there is not much difference in the compensation effect when siblings are further separated in age.

In Row 2, we re-estimate our model without the covariates (i.e., maternal age, birth order, birth year, birth quarter, HAZ, birth place, and the type of sibling), only controlling for age: the IV-FE estimate equals -0.045 (z=-2.81). This result provides extra support for our assumption that rainfall is exogenously determined because it shows that our estimate is not conditional on the set of control variable included in the model.

In Row 3, we show results using only the private tuition fees as the dependent variable and find consistent results, which are higher in magnitude. When parents observe an increase of one standard deviation in ability, they reduce private tuition fees by 9.9% of a standard deviation. Next we investigate whether the likelihood to take private tuition is contingent upon cognitive ability, using a dummy variable of taking private tuition as the dependent variable. Shown in Row 4, the result suggests a compensating behaviour of parents: the probability of offering private tuition to a child will increase by 30.8% if the child is under developed by one standard deviation in cognitive ability.

We also exploit child time use as a potential measure for educational investment. Firstly, we study parental responses using study hours. Row 5 shows that the coefficient is negative and significant, with the estimate equals -0.542 (z=-2.16), suggesting a child's study hours will increase by 54.2% of standard deviation if the child's ability is one standard deviation lower. In Row 6, we

TABLE 2.6: Robustness regression models

Model Variations	Obs	First-stage: Average rainfall from 0-24 months on ability	Second-stage: Ability on parental responses
(1) Siblings born at least three years apart	844	0.187***	-0.036***
		(0.033)	(0.014)
			[-0.062,-0.014]
(2) Only control for age	1402	0.143***	-0.045***
		(0.027)	(0.016)
			[-0.078,-0.022]
(3) Private tuition fees as outcome	1402	0.156***	-0.099***
		(0.027)	(0.036)
			[-0.167,-0.045]
(4) Having private tuition (dv) as outcome	1402	0.156***	-0.308***
		(0.027)	(0.085)
			[-0.480,-0.187]
(5) Study hours at home as outcome	1402	0.156***	-0.542**
		(0.027)	(0.251)
			[-1.000,-0.150]
(6) Care, chore, task, work hours as outcome	1402	0.156***	0.425*
		(0.027)	(0.254)
			[0.029, 0.888]
(7) Siblings born at the same place	1392	0.156***	-0.038***
		(0.027)	(0.013)
			[-0.064,-0.019]
(8) Siblings both younger than 8 in R3	1220	0.164***	-0.040***
		(0.029)	(0.013)
			[-0.065,-0.021]
(9) Siblings both enrolled	1170	0.155***	-0.038***
		(0.027)	(0.013)
			[-0.064,-0.019]
(10) Not control for HAZ	1402	0.165***	-0.039***
		(0.029)	(0.013)
			[-0.065, -0.019]
(11) Rainfall on consumer index	1402	-0.005	
		(0.009)	

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Model 1 to 10 use the same IV-FE model as the one for the main result in Table 2.4, instrumenting the ability using average rainfall during the first two years of life, whilst Model 11 uses cross-section OLS estimation. Robust standard errors in parentheses. The weak IV robust AR confidence intervals are in the brackets in column 3. Row 1 uses a sub-sample which contains the sibling-pairs which have an age gap of at least 3 years. Row 2 removes all the covariates displayed in Table 2.4 apart from age of child. Row 3 uses private tuition fees as dependent variable. Row 4 uses whether the child receives private tuition as the dummy outcome variable. Row 5 exploits child study hours at home as outcome variable. Row 6 uses summary of hours spend by child in care, chore, task and work as outcome variable. Row 7 restricts the sample to siblings born at the same place. Row 8 restricts the sample to siblings both younger than 8 in Round 3. Row 9 restricts the sample to siblings both enrolled. Row 10 removes height-for-age Z-score as control in the model. Row 11 is the OLS estimation regressing consumer durable index on the average rainfall from 0 to 24 months.

examine parental responses in respect to child hours spent in caring, doing chores and tasks, and working and find the coefficient being a positive and significant, with the estimate equals 0.425 (z=1.68), suggesting a child might spend more time in care, chores, tasks and work by 43.5% of standard deviation when the child's ability is higher by one standard deviation. This implies consistent compensating parental responses in terms of child time use.

We restrict the sample to siblings born at the same place and find no change in the coefficient, with the estimate equals to -0.038 (z=-2.92), as shown in Row 7. In Row 8, when the sample is restricted to children younger than eight (i.e., the index child is paired with a young sibling), the coefficient does not change much, with the estimate equal to -0.040 (z=-3.08). In Row 9, the compensating effect is consistent using a sub-sample of siblings both enrolled in school, as the estimate equals to -0.038 (z=-2.92). In Row 10, we drop the control of HAZ in the specification, and it does not change the result; the estimate is -0.039 (z=-3.00).

In Row 11, we regress the household consumer durables index in Round 4 on rainfall in early life and find no significant effect (t=0.56), implying that rainfall shocks in early life do not have persistent impact on consumption pattern of the household in the later life of children. It supports our assumption that rainfall in early life does not affect parental investment in later life through a direct mechanism.

Finally, we checked the difference in increment of outcomes of the two siblings between Round 5 (2016) and Round 4 (2013), to examine whether the investment differential did close the gap in ability. In a difference-in-difference specification, we found the coefficient on investment was negative and statistically insignificant, with a p-value of 0.24. So, at least in the three year period, attempts to compensate were largely unsuccessful which may be i) due to the relatively small magnitude of the difference in investments (3.9% of a standard deviation at the mean), or ii) because the low level of early ability constrains the return to later investment, consistent with Cunha and Heckman (2007)'s hypothesis of *dynamic complementarities* in the human capital production function.

2.6 Conclusion

We find that for a sample of poor Ethiopian households, on average parental investment compensates weakly for a low-ability outcome. We use an instrumental variable approach combined with panel data and a sibling fixed-effects model to provide robust evidence. This is of policy relevance since the results suggest that the detrimental effects of early life shocks might be mediated or muted by parental responses and hence the biological effects of early nutritional shocks might be larger than policymakers observe. In addition, it complements the literature on reduced-form estimates of the total effect of an early life shock or adverse event on final adult health in Ethiopia (e.g. Dercon and Porter (2014)).

This finding is in line with results from some previous studies reporting compensating parental behaviour (Behrman, 1988; Bharadwaj et al., 2018; Rosenzweig and Wolpin, 1988; Del Bono et al., 2012; Frijters et al., 2009; Halla and Zweimüller, 2014; Leight, 2017; Yi et al., 2015). It is also consistent with the intrafamily resource allocation model introduced by Behrman et al. (1982), suggesting parents favour equity over efficiency.

However, we have indicative evidence that this effect varies across family SES. Relatively advantaged parents provide more cognitive investment to lower-ability children, and lower-class families exhibit only small and modest compensatory behaviours. The finding is consistent across all measures of parental socio-economic advantage (maternal education, household wealth and household size), though the 95% confidence intervals for the estimates overlap. Consistent with prior findings, mothers with higher education compensate for lower-endowed children (Cabrera-Hernandez, 2016; Hsin, 2012; Restrepo, 2016).

Our results therefore complement the literature which studies whether the effect of shocks to early ability can be eliminated or mitigated through investments, that themselves depend on family socio-economic status. Most studies have found that compared with the low-ability children born in higher-class families, the low-ability child born in lower-class families have worse outcomes in adulthood. One hypothesis of the results in the literature is that parental involvement plays a role in reinforcing the poor ability outcome. Specifically, higher-class parents compensate for the differences in ability, or at least are not reinforcing the differences. Our results support the hypothesis that parental investment varies by family SES, even in a context of low-income by international standards. What is difficult to disentangle, given the high correlation between SES as measured by wealth, and by parental education, is to differentiate whether high-SES parents are more able to observe the difference in ability; more able to compensate for the difference; or both of these. More work on this issue is needed where suitable data can be collected.

Bibliography

Abufhele, Alejandra, Jere Behrman, and David Bravo (2017). “Parental preferences and allocations of investments in children’s learning and health within families”. *Social Science & Medicine* 194, pp. 76–86.

Adhvaryu, Achyuta and Anant Nyshadham (2016). “Endowments at birth and parents’ investments in children”. *The Economic Journal* 126.593, pp. 781–820.

Aizer, Anna and Flávio Cunha (2012). *The production of human capital: Endowments, investments and fertility*. Working Paper 18429. National Bureau of Economic Research.

Almond, Douglas and Janet Currie (2011). “Killing me softly: The fetal origins hypothesis”. *The Journal of Economic Perspectives* 25.3, pp. 153–172.

Almond, Douglas and Bhashkar Mazumder (2013). “Fetal origins and parental responses”. *Annual Review of Economics* 5.1, pp. 37–56.

Almond, Douglas, Janet Currie, and Valentina Duque (2018). “Childhood circumstances and adult outcomes: Act II”. *Journal of Economic Literature* 56.4, pp. 1360–1446.

Angrist, Joshua D and Jörn-Steffen Pischke (2008). *Mostly harmless econometrics: An empiricist’s companion*. Princeton University Press.

Ayalew, Tekabe (2005). “Parental preference, heterogeneity, and human capital inequality”. *Economic Development and Cultural Change* 53.2, pp. 381–407.

Barnett, Inka et al. (2012). “Cohort profile: The Young Lives study”. *International Journal of Epidemiology* 42.3, pp. 701–708.

Becker, Gary S (1991). *A Treatise on the Family*. Harvard University Press.

Becker, Gary S and Nigel Tomes (1976). “Child endowments and the quantity and quality of children”. *Journal of Political Economy* 84.4, Part 2, S143–S162.

Behrman, Jere R (1988). "Nutrition, health, birth order and seasonality: Intra-household allocation among children in rural India". *Journal of Development Economics* 28.1, pp. 43–62.

Behrman, Jere R, Robert A Pollak, and Paul Taubman (1982). "Parental preferences and provision for progeny". *Journal of Political Economy* 90.1, pp. 52–73.

Behrman, Jere R, Mark R Rosenzweig, and Paul Taubman (1994). "Endowments and the allocation of schooling in the family and in the marriage market: The twins experiment". *Journal of Political Economy* 102.6, pp. 1131–1174.

Bhalotra, Sonia R and Damian Clarke (2018). "The twin instrument: Fertility and human capital investment". *IZA Institute of Labor Economics Discussion Paper* 11878.

Bharadwaj, P, J Eberhard, and C Neilson (2018). "Health at birth, parental investments, and academic outcomes". *Journal of Labour Economics* 36.2.

Bound, John and Gary Solon (1999). "Double trouble: On the value of twins-based estimation of the return to schooling". *Economics of Education Review* 18.2, pp. 169–182.

Buckles, Kasey and Shawna Kolka (2014). "Prenatal investments, breastfeeding, and birth order". *Social Science & Medicine* 118, pp. 66–70.

Cabrera-Hernandez, Francisco (2016). "The accident of birth: Effects of birthweight on educational attainment and parents' compensations among siblings". *Center for Economic Research and Teaching (CIDE)*.

Campbell, Jonathan (1998). "Book Review: Peabody Picture Vocabulary Test". *Journal of Psychoeducational Assessment* 16.4, pp. 334–338.

Casale, Daniela and Chris Desmond (2016). "Recovery from stunting and cognitive outcomes in young children: evidence from the South African Birth to Twenty Cohort Study". *Journal of Developmental Origins of Health and Disease* 7.2, pp. 163–171.

Clarke, Damian et al. (2017). "PLAUSEXOG: Stata Module to implement Conley et al's Plausibly Exogenous Bounds". *Statistical Software Components*.

Conley, Dalton (2008). "Bringing sibling differences in: Enlarging our understanding of the transmission of advantage in families". *Social Class: How Does It Work?*, pp. 179–200.

Conley, Timothy G, Christian B Hansen, and Peter E Rossi (2012). "Plausibly Exogenous". *Review of Economics and Statistics* 94.1, pp. 260–272.

Crookston, Benjamin T et al. (2013). "Postinfancy growth, schooling, and cognitive achievement: Young Lives". *The American Journal of Clinical Nutrition* 98.6, pp. 1555–1563.

Cueto, Santiago and Juan Leon (2012). *Psychometric characteristics of cognitive development and achievement instruments in round 3 of Young Lives*. Tech. rep. 15. Young Lives.

Cunha, Flavio and James Heckman (2007). "The technology of skill formation". *American Economic Review* 97.2, pp. 31–47.

Currie, Janet and Douglas Almond (2011). "Human capital development before age five". *Handbook of Labor Economics*. Vol. 4. Elsevier, pp. 1315–1486.

Datar, Ashlesha, M. Rebecca Kilburn, and David S. Loughran (2010). "Health endowments and parental investments in infancy and early childhood". *Demography* 47.6, pp. 145–162.

Del Bono, Emilia, John Ermisch, and Marco Francesconi (2012). "Intrafamily resource allocations: a dynamic structural model of birth weight". *Journal of Labor Economics* 30.3, pp. 657–706.

Dendir, Seife (2014). "Children's cognitive ability, schooling and work: Evidence from Ethiopia". *International Journal of Educational Development* 38, pp. 22–36.

Dercon, Stefan and Catherine Porter (2014). "Live aid revisited: Long-term impacts of the 1984 Ethiopian famine on children". *Journal of the European Economic Association* 12.4, pp. 927–948.

Doyle, Orla, Colm P Harmon, James J Heckman, and Richard E Tremblay (2009). "Investing in early human development: Timing and economic efficiency". *Economics & Human Biology* 7.1, pp. 1–6.

Dunn, Lloyd M and Leota M Dunn (1997). *PPVT-III: Peabody picture vocabulary test*. American Guidance Service.

Finlay, Keith, Leandro Magnusson, Mark E Schaffer, et al. (2016). "WEAKIV: Stata module to perform weak-instrument-robust tests and confidence intervals for instrumental-variable (IV) estimation of linear, probit and tobit models". *Statistical Software Components*.

Frijters, Paul, David W. Johnston, Manisha Shah, and Michael A. Shields (2009). "To work or not to work? Child development and maternal labor supply". *American Economic Journal: Applied Economics* 1.3, pp. 97–110.

Frijters, Paul, David W Johnston, Manisha Shah, and Michael A Shields (2013). "Intrahousehold resource allocation: Do parents reduce or reinforce child ability gaps?" *Demography* 50.6, pp. 2187–2208.

Glewwe, Paul, Hanan G Jacoby, and Elizabeth M King (2001). "Early childhood nutrition and academic achievement: A longitudinal analysis". *Journal of Public Economics* 81.3, pp. 345–368.

Grätz, Michael and Florencia Torche (2016). "Compensation or Reinforcement? The stratification of parental responses to children's early ability". *Demography* 53.6, pp. 1883–1904.

Griliches, Zvi (1979). "Sibling models and data in economics: Beginnings of a survey". *Journal of Political Economy* 87.5, Part 2, S37–S64.

Halla, Martin, Martina Zweimüller, et al. (2014). *Parental response to early human capital shocks: Evidence from the Chernobyl Accident*. Tech. rep. The Austrian

Center for Labor Economics and the Analysis of the Welfare State, Johannes Kepler University Linz, Austria.

Heckman, James J (2007). "The economics, technology, and neuroscience of human capability formation". *Proceedings of the National Academy of Sciences of the United States of America*, pp. 13250–13255.

Hill, Ruth Vargas and Catherine Porter (2017). "Vulnerability to drought and food price shocks: Evidence from Ethiopia". *World Development* 96, pp. 65–77.

Hsin, Amy (2012). "Is biology destiny? Birth weight and differential parental treatment". *Demography* 49.4, pp. 1385–1405.

Kleibergen, Frank and Richard Paap (2006). "Generalized reduced rank tests using the singular value decomposition". *Journal of Econometrics* 133.1, pp. 97–126.

Lareau, Annette (2011). *Unequal childhoods: Class, race, and family life*. Univ of California Press.

Leight, Jessica (2017). "Sibling rivalry: Endowment and intrahousehold allocation in Gansu Province, China". *Economic Development and Cultural Change*, pp. 456–493.

Leon, Juan and Abhijeet Singh (2017). *Equating test scores for receptive vocabulary across rounds and cohorts in Ethiopia, India and Vietnam*. Tech. rep. Young Lives.

Lynch, Jamie L and Ryan Brooks (2013). "Low birth weight and parental investment: Do parents favor the fittest child?" *Journal of Marriage and Family* 75.3, pp. 533–543.

Maccini, Sharon and Dean Yang (2009). "Under the weather: Health, schooling, and economic consequences of early-life rainfall". *American Economic Review* 99.3, pp. 1006–26.

Nicoletti, Cheti, Kjell G Salvanes, and Emma Tominey (2018). "Response of parental investments to child's health endowment at birth". *Health Econometrics*, p. 175.

Olea, José Luis Montiel and Carolin Pflueger (2013). "A robust test for weak instruments". *Journal of Business & Economic Statistics* 31.3, pp. 358–369.

Oumer, Jeilu (2009). *The challenges of free primary education in Ethiopia*. UNESCO. International Institute for Educational Planning.

Outes-Leon, Ingo and Alan Sanchez (2008). *An assessment of the Young Lives sampling approach in Ethiopia*. Tech. rep. 1. Young Lives.

Paxson, Christina and Norbert Schady (2007). "Cognitive development among young children in Ecuador the roles of wealth, health, and parenting". *Journal of Human Resources* 42.1, pp. 49–84.

Pflueger, Carolin and Su Wang (2015). "A robust test for weak instruments in Stata". *Stata Journal* 15.1, pp. 216–225.

Price, Joseph (2008). "Parent-child quality time: Does birth order matter?" *Journal of Human Resources* 43.1, pp. 240–265.

Restrepo, Brandon J (2016). "Parental investment responses to a low birth weight outcome:Who compensates and who reinforces?" *Journal of Population Economics* 29.4, pp. 969–989.

Rosales-Rueda, Maria Fernanda (2014). "Family investment responses to childhood health conditions: Intrafamily allocation of resources". *Journal of Health Economics* 37, pp. 41–57.

Rosenzweig, Mark R and Kenneth I Wolpin (1988). "Heterogeneity, intrafamily distribution, and child health". *Journal of Human Resources*, pp. 437–461.

Schaffer, Mark E et al. (2015). "Xtivreg2: Stata Module to perform extended IV/2SLS, GMM and AC/HAC, LIML and k-class regression for panel data models". *Statistical Software Components*.

Staiger, Douglas, James H Stock, et al. (1997). "Instrumental variables regression with weak instruments". *Econometrica* 65.3, pp. 557–586.

UNICEF (2009). "Abolishing school fees in Africa lessons from Ethiopia, Ghana, Kenya, Malawi, and Mozambique".

Victora, Cesar Gomes, Mercedes de Onis, Pedro Curi Hallal, Monika Blössner, and Roger Shrimpton (2010). "Worldwide timing of growth faltering: Revisiting implications for interventions". *Pediatrics*, peds–2009.

Yi, Junjian, James J Heckman, Junsen Zhang, and Gabriella Conti (2015). "Early health shocks, intra-household resource allocation and child outcomes". *The Economic Journal* 125.588, F347–F371.

Appendix

TABLE A1: Full list of coefficients from Table 4

Dependent variable: total educational fees	IV-FE result
Cognitive ability	-0.038**
	(0.017)
First born	-0.484***
	(0.107)
Second born	-0.441***
	(0.096)
Third born	-0.395***
	(0.083)
Forth born	-0.343***
	(0.071)
Fifth born	-0.281***
	(0.058)
Sixth born	-0.232***
	(0.046)
Seventh born	-0.162***
	(0.032)
Eighth born	-0.115***
	(0.021)
Ninth born	-0.054***
	(0.012)
Old sister with young brother	0.060***
	(0.015)
Old brother with young sister	0.054***
	(0.015)
Old brother with young brother	0.069***
	(0.017)
Old sister with young sister	0.066***
	(0.016)
HAZ in Round 3	0.003
	(0.002)
Born in first quarter	-0.005
	(0.017)
Born in second quarter	-0.005
	(0.013)
Born in third quarter	0.001
	(0.008)
Age in month	0.012
	(0.017)
Square of age in month	-0.000
	(0.000)
Cubic of age in month	0.000
	(0.000)
Maternal age at birth	0.000
	(0.005)
Constant	-0.017*
	(0.009)

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Full list of main result in column 3 of Table 2.4, instrumenting the ability using average rainfall during the first two years of life. Community clustered standard errors in parentheses. Birth order's reference group is tenth born. Birth quarter's reference group is born in forth quarter. Birth year and birthplace are not shown in this table, but controlled in the regressions.

TABLE A2: Robustness check: First stage results adding sibling rainfall in infancy using community fixed-effects model

COGNITIVE ABILITY	(1)	(2)
Child average rainfall in the first two years of life	0.124	0.104
	(0.032)***	(0.043)**
Sibling average rainfall in the first two years of life		-0.027
		(0.041)
Child-level controls	Yes	Yes
Household-level controls	Yes	Yes
Mother-level controls	Yes	Yes
Community fixed-effects	Yes	Yes
Observations	1402	1402
R-squared	0.674	0.674

Note:* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are in parentheses. This table reports analogous regression as the one in the first-stage regression. The dependent variable is cognitive ability. Children controls are age in months, square of age in months, cubic of age in months, maternal age at birth, birth order, birthplace, birth quarter, birth year, language, and the type of sibling (such as born as an older sister and paired with a younger brother). Household-level controls are household size, wealth index, and gender of household head. Mother-level controls are a series of levels of maternal education.

TABLE A3: Redundancy tests: Cognitive ability and cognitive resources

DEPENDENT VARIABLE:	IV-FE	
TOTAL EDUCATIONAL FEES	Instruments: Rainfall from *in utero* to year 2	Instruments: Rainfall at birth & in year 1
Cognitive ability	-0.033***	-0.039***
	(0.011)	(0.013)
Weak identification test:		
Moutiel-Pflueger effective F stat	11.025	17.169
Moutiel-Pflueger critical values:		
5% of worst case bias	21.195	5.808
10% of worst case bias	12.773	4.550
20% of worst case bias	8.182	3.828
IV redundancy test:		
Redundancy of rainfall *in utero* p-val	0.270	-
Redundancy of rainfall at birth p-val	0.000	0.003
Redundancy of rainfall in year 1 p-val	0.006	0.005
Redundancy of rainfall in year 2 p-val	0.007	-
Observations	1402	1402
No. Excluded Instruments	4	2

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. Children controls are age, maternal age at birth,height-for-age Z-score, birthplace, birth quarter, birth order, birth year, and the type of sibling. IV redundancy test is a LM test of a specified instrument, asking whether this instrument provides useful information to identify the equation. The null hypothesis is the instrument does not contribute to the asymptotic efficiency of the equation. Rejecting the null suggests that the specified instrument does capture information of the endogenous variable.

TABLE A4: Robustness check: First stage results using restricted sample

Cognitive Ability	(1)	(2)	(3)	(4)	(5)	(6)
Rainfall *in Utero*	0.020					
	(0.043)					
	[0.037]					
Rainfall at Birth		0.122				0.075
		(0.024)***				(0.032)**
		[0.030]***				[0.039]*
Rainfall in Year 1			0.206			0.128
			(0.040)***			(0.053)**
			[0.050]***			[0.067]*
Rainfall in Year 2				-0.013		
				(0.046)		
				[0.071]		
Average Rain at Birth and in Year 1					0.187	
					(0.033)***	
					[0.040]***	
Weak instrument test:						
Montiel-Pflueger effective F stat	0.219	25.092	26.370	0.076	32.859	15.586
Montiel-Pflueger critical values:						
5% of worst case bias	37.418	37.418	37.418	37.418	37.418	7.119
10% of worst case bias	23.109	23.109	23.109	23.109	23.109	5.315
20% of worst case bias	15.062	15.062	15.062	15.062	15.062	4.267
Observations	844	844	844	844	844	844

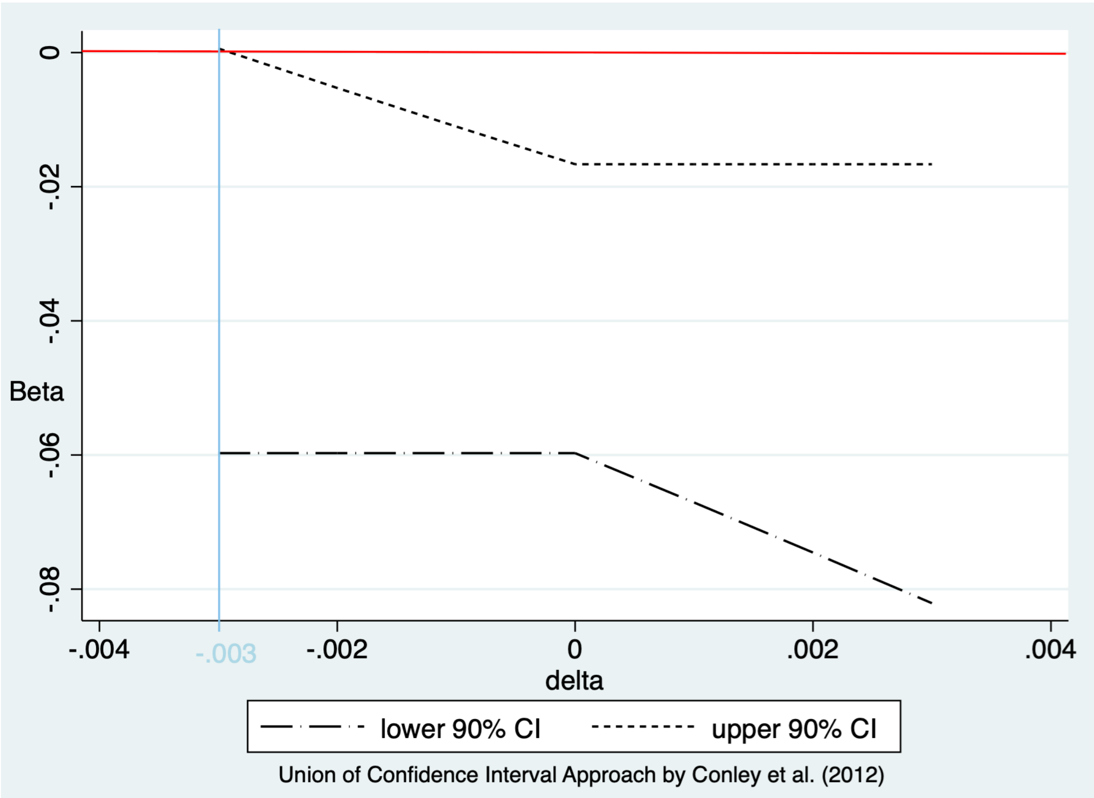
Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. Clustered standard errors by community are in brackets. These are within-household fixed effects estimates from the first stage of the IV-FE model, same as the one in Table 2.3. This sub-sample contains the sibling-pairs which have an age gap of at least 3 years. Controls are age, maternal age at birth, birth order, birthplace, birth quarter, birth year, height-for-age in Round3 and the type of sibling (such as born as an older sister and paired with a younger brother).

TABLE A5: IV-FE results: HAZ as the interest variable

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: HAZ	**Panel A: First-stage IV regressions on Height-for-age Z-score**					
Rainfall *in Utero*	-0.047					
	(0.049)					
	[0.041]					
Rainfall at Birth		-0.018				-0.041
		(0.034)				(0.040)
		[0.030]				[0.042]
Rainfall in Year 1			0.033			0.064
			(0.047)			(0.054)
			[0.049]			[0.066]
Rainfall in Year 2				-0.007		
				(0.064)		
				[0.066]		
Average Rain at Birth and Year 1					-0.003	
					(0.046)	
					[0.039]	
Weak IV test: MP F stat	0.950	0.288	0.493	0.010	0.003	0.836
Dependent variable: total educational fees	**Panel B: Second-stage IV regressions on total educational fees**					
	Instruments: Rainfall in *Utero*	Instruments: Rainfall at birth	Instruments: Rainfall in year 1	Instruments: Rainfall in year 2	Instruments: Average rainfall in the first two years of life	Instruments: Rainfall & at birth rainfall in year 1
Height-for-age Z-score	-0.012	0.187	-0.195	-0.196	2.248	-0.030
	(0.044)	(0.350)	(0.273)	(1.865)	(37.914)	(0.038)
	[0.040]	[0.300]	[0.275]	[1.933]	(32.260)	(0.034)
Observations	1402	1402	1402	1402	1402	1402

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. Clustered standard errors by community in brackets. Child controls include age in months, square of age in months, cubic of age in months, maternal age at birth, birth order, birthplace, birth quarter, birth year, and the type of sibling (such as born as an older sister and paired with a younger brother). The weak instrument test is robust to heteroskedasticity. The Montiel-Pflueger (MP) F statistics are very similar to Kleibergen-Paap rk Wald F statistics in weak instrument test. The MP weak instrument test offers valid critical values at 95% confidence level and test statistics in the absence of assumption of i.i.d. data. The critical values are the same as the ones in Table 2.3.

FIGURE B1: Estimated β by direct effect of instrument



Chapter 3

Is maternal autonomy a potential pathway to child nutritional status? Evidence from Ethiopia 2005-2016

[1]

[1]The study uses Ethiopia Demographic and Health Survey (EDHS) data, 2005 to 2016. The funding for the EDHS was provided by the government of Ethiopia, the United States Agency for International Development (USAID), the government of the Netherlands, the Global Fund, Irish Aid, the World Bank, the United Nations Population Fund (UNFPA), the United Nations Children's Fund (UNICEF), and UN Women. ICF provided technical assistance through the DHS Programme, a USAID-funded project providing support and technical assistance in the implementation of population and health surveys. The views expressed here are those of the author(s). They are not necessarily those of EDHS or other funders. Thanks to Achim Ahrens, Ellen Alem, Martha Kibur, Margaret Leighton, Alula Pankhurst, Remy Pigois, Mark Schaffer, Yisak Tafere, Emma Tominey, and Vincenzo Vinci for helpful suggestions. Thanks to seminar and conference participants in UNICEF Ethiopia, Jun 2018; Adam Smith Development Economics workshop, Scotland, Dec 2018; SGPE conference, Scotland, Jan 2019.

Abstract

A small but increasing body of literature finds a correlation between women's empowerment and child nutritional status, but the causal evidence is limited, given the challenges to identification. This study provides evidence on the impact of maternal autonomy on child nutrition in Ethiopia, using three rounds of Demographic and Health Survey data from 2005 to 2016. To deal with identification issues raised by omitted variables bias, a novel method, the post-double-selection Lasso by Belloni et al. (2014) is combined with village-wave fixed effects. This allows us to select high-dimensional raw regressors that have strong predictive power on maternal autonomy and child nutrition without concerns about having a sparse model. The findings indicate that one standard deviation increase in maternal autonomy can reduce the probability of a child being underweight by 2%, and that this effect is most apparent in the older group of children aged 24-59 months who are most likely weaned. The effect size is moderate yet significant. However, no impact is found on stunting. The study sheds light on the importance of improving maternal autonomy for children in the critical developmental period to prevent adverse outcomes of children, especially when children are weaned.

Keywords: Child Nutrition, Maternal Autonomy, Bayesian Confirmatory Factor Analysis, Post-double-selection Lasso.

JEL classification: I1, J13, J16, O15.

3.1 Introduction

Since the 20th century, gender inequality has become one of the leading political concerns in developing countries. Women's empowerment has been specifically included as one of the main goals in Millennium Development Goals (MDG) and Sustainable Development Goals (SDG). For example, MDG goal 3 and SDG goal 5 have been exclusively targeted at achieving gender equality and empowering all women and girls. The impetus to empower women in developing countries is twofold: firstly, for its own merit, to reduce the disparity of power between women and men under a wider frame of social justice; secondly, to benefit child health (Duflo, 2003; Lépine and Strobl, 2013; Maitra, 2004). A systematic review of 67 eligible quantitative studies showed a positive association between women's empowerment and child health outcomes in the developing world (Pratley, 2016).

In low income countries, it is not only that the high level of nutrition deprivation among young children is of profound concern but also the slow progress of improvements in nutritional indicators. For instance, Ethiopia is one of the most undernourished populations in the world, where over a half of preschool children are stunted and almost a third are severely stunted[2] in 2000, while stunting prevalence reduced only by 1.4% each year till 2011 (Headey, 2014). More worryingly, malnutrition triggers nearly a half of under-five deaths in Ethiopia.[3]

Though the literature has identified a positive relationship between women's empowerment and child nutrition, some limitations have been brought into

[2]The World Health Organisation (WHO) defines a child health status by comparing the height to the standards developed using data collected in the WHO Multicentre Growth Reference Study(WHO, 2006). A child is identified as stunted when a Height-for-Age Z-score (HAZ) is less than -2, and a child is defined as severely stunted when a HAZ is less than -3 (WHO, 2010).

[3]WHO website, Ethiopia Child Health page. Access at http://www.afro.who.int/health-topics/health-topics-ethiopia.

the discussion and carefully reviewed (Carlson et al., 2015; Lépine and Strobl, 2013; Pratley, 2016; Richardson, 2018). First, the measures adopted in many studies are indirect proxies of women's empowerment, such as income earned by women and asset ownership, which might reflect other economic aspects of a household, and may not necessarily capture the component of autonomy. In consequence, more studies have zeroed in on the direct evidence of empowerment (Kishor, 2000; Richardson, 2018) by relying on specific (questionnaire-based) measures of autonomy. Second, many studies construct a proxy for autonomy using a simple additive approach, summarising questionnaire responses with an assumption of equal importance for each item, and thus neglecting the multi-dimensional construct of autonomy. And last but not least, the majority of the existing literature fails to deal with the potential endogeneity of autonomy even though a direct measure of autonomy is employed, since unobservables correlated with both autonomy and child outcomes might bias the estimates of the effects.

The present study contributes to the literature in three ways. Firstly, we are the first to provide comprehensive evidence on the causal relationship between women's autonomy and child nutritional status in Ethiopia, using a nationally representative survey, the Ethiopian Demographic and Health Surveys (EDHS). Secondly, we adopt a holistic concept of autonomy, hypothesising that higher level of maternal autonomy shapes positive intrinsic-motivated maternal behaviours which enhance child nutritional status (Shroff et al., 2011). Following the recent literature, we measure autonomy as a multi-dimensional construct, since autonomy can be manifested over correlated but independent sub-components which have been established by the literature, including household decision-making, non-acceptance of domestic violence, and control over sexual behaviours (Agarwala and Lynch, 2006; Sandberg and Rafail, 2013; Shroff et al., 2011). Therefore, we acknowledge the construct of autonomy by measuring it using a reliable statistical framework based on the theory of autonomy, i.e. Confirmatory Factor Analysis (CFA). To allow for looser assumptions of the model and more accurate measure, we innovatively use a Bayesian CFA model. Thirdly, to deal with potential endogeneity caused by omitted variables, we take advantage of an innovative "post-double-selection Lasso" (PDS-LASSO) method (Belloni et al., 2014) combined with village-wave fixed effects. A pressing problem in the development literature is that researchers might choose control variables consciously or unconsciously to produce an estimate with statistical significance and expected signs. The PDS-LASSO method enables us to tackle this problem and shows that the results are robust, even when we use raw regressors available from the data, interactions and transformations of these to proxy for inputs of nutrition, without concerns of a sparse model. To preview the results, we find that improved women's autonomy can induce better nutritional status of children in Ethiopia. Specifically, one standard deviation increase in maternal autonomy can reduce the probability of a child being underweight by 2%, yet no impact is found on stunting. This is mainly found among children between two and five years old, i.e. when children are weaned. Although the coefficient is small, it is significant, as the baseline of the prevalence of underweight is still as high as 18% in 2016 in Ethiopia.

The remainder of the paper is organised as follows. In the next section, we briefly present the relevant literature. In the third section, data and the construction of variables are described, followed by our econometric method, results, robustness check, and concluding remarks.

3.2 Literature review

3.2.1 Definition and measurement of women's autonomy

The term "women's empowerment" has been widely used to depict a broad range of concepts and has been utilised to advocate for political promotion to improve women's status (Malhotra et al., 2005). The most influential definition of women's empowerment provided by Kabeer (1999), Malhotra et al. (2005) reinforces the definition as the development of women's ability to make decisions for their lives and living conditions. Malhotra et al. (2005)'s definition emphasises women's agency as the ability to make choices and allow oneself to exercise power, during which women attain agency over time. It has been agreed that when using cross-sectional data for analysis, agency can be referred to as autonomy to distinguish a static state from a process, dropping the element of the temporal variations in the initial definition (Allendorf, 2012; Malhotra et al., 2005; Jejeebhoy and Sathar, 2001). In Dyson and Moore (1983)'s classic paper, autonomy has been defined as "the ability ... to obtain information and to use it as the basis for making decisions about one's private concerns and those of one's intimates" (p.45). In respect of the family-level autonomy, it reveals a woman's ability to attain information, make use of it, and therefore have control over decisions on both personal and family levels (Carlson et al., 2015).

As such an elusive and abstract concept is not easy to measure and poses methodological challenges, Richardson (2018) recommends that researchers who attempt to measure women's empowerment should use direct indicators of empowerment, i.e. autonomy (agency), when possible. In a cross-sectional setting, using indirect indicators, such as education and ownership of land, might provide inadequate evidence (Malhotra and Mather, 1997; Jejeebhoy and Sathar, 2001). As these are "resources" that could enhance women's agency, they are also correlated with overall wealth and socioeconomic status, and are therefore not necessarily informative (Kabeer, 1999; Richardson, 2018). Therefore, increasing studies have started to use autonomy as a measure of empowerment.

It is widely acknowledged that autonomy is a multi-dimensional construct (Agarwala and Lynch, 2006; Malhotra et al., 2005), which likely consists of correlated but independent sub-components (Gupta and Yesudian, 2006; Mason and Smith, 2000). For example, a woman in some Indian societies with high autonomy in making household decisions does not necessarily have the freedom to travel alone (Gupta and Yesudian, 2006). In the recent literature, survey data is the main instrument for measuring autonomy. A widely used example is the Demographic and Health Surveys (DHS), in which N different dimensions are included, such as household decision-making and non-acceptance of domestic violence. To parse the distinctive components, Confirmatory Factor Analysis (CFA) has also been used to identify the dimensional construct of autonomy based on a weighted method, rather than a simple additive approach (Agarwala and Lynch, 2006; Pratley and Sandberg, 2018; Sandberg and Rafail, 2013; Shroff et al., 2011).

Apart from using an additive approach to measure autonomy, another limitation for a large body of previous studies, especially the ones using DHS data, is that they measure autonomy based on items that reflect responsibilities rather than autonomy, as has been argued by Basu and Koolwal (2005). The questionable items often include measures about whether one can make meals, whether one is free to go to the market, and whether a woman can take a sick child to the hospital. In countries where taking care of children, cooking and buying food are female's responsibilities, and thus such activities are often well-supported, these items do not reveal one's freedom of social activities and exercising power. In consequence, it is argued that the positive correlation found between these indicators and child outcome might not indicate that the woman is a free agent, which contributes to a better outcome of a child, but merely an instrumental effect of mother's fulfilling the responsibility of being a good mother in an orthodox gender role (Basu and Koolwal, 2005; Iversen and Palmer-Jones, 2018).

This is supported by the well established Self Determination Theory from psychology (Deci and Ryan, 2000), where autonomy is conceptualised as one of the three innate psychological needs for intrinsic motivation to stimulate positive actions. Psychologists argue that the *sense of autonomy*, or in attributional terms, internal perceived locus of causality (DeCharms, 1968), is essential for people to maintain intrinsic motivation. However, it is noted that a person's actions can be the product of intrinsic motivation, extrinsic motivation (i.e. gender norms), or both (Deci and Ryan, 2000; Seymour and Peterman, 2018; Vaz et al., 2016). Therefore, Seymour and Peterman (2018) raise the issue that using some decision-making indicators behind which the motivation is extrinsic to proxy for autonomy could be problematic. For example, a woman taking care of children might be a result of fears of disapproval from her husband and community (Seymour and Peterman, 2018).[4] Therefore, we follow Basu and Koolwal (2005)'s paper to select carefully among those measures that are more robust in the DHS which represent women's self-interest and their private concerns, such as decision on visits to relatives, decision on self health care, decision on using husband's earnings, decision on purchasing large assets, control over sex and non-acceptance of violence.

[4]The indicators of making meals, taking a sick child to the hospital and free to go to the market is incidentally unavailable in EDHS.

3.2.2 The relationship between women's autonomy and child's nutrition

The mechanism of how maternal autonomy impacts on child health might lie in the occurrence of a willing and positive behaviour change in child care resulted by mother's intrinsically motivational autonomy, following the Self Determination Theory (Shroff et al., 2011). The self-motivational autonomy stimulates positive child care behaviours and enhances the reflection of a woman's values through her actions, which can bring changes in child health and wellbeing. Meanwhile, scholars have agreed that women may have a stronger preference for child health than men (Castle, 1993; Duflo, 2003; Schmeer, 2005; Thomas, 1990).

Although women's autonomy has been studied in explaining various outcomes, such as child immunisation (Malhotra et al., 2014; Thorpe et al., 2016), the effect of maternal autonomy on child nutritional outcomes is understudied, and the majority of the evidence is from biological studies. To the best of our knowledge, only a few paper study the association between maternal autonomy and child nutrition, and the findings are not always consistent due to the following reasons: (1) what has been used for measuring autonomy is not consistent; (2) the measurement of child nutrition varies; and (3) the effect may be found in specific age groups.

Using a small sample from a northern Kenyan group, Brunson et al. (2009) define maternal autonomy using women's ability to make decisions, to control over her body, and to determine how to use resources. A composite score for overall autonomy is constructed using Cronbach's alpha, and nutrition is proxied by a short-term measure, Weight-for-Height Z-score. While no significant relationship is found for children in the age group 0-36 months, there is a significant and positive relationship for the older group aged 3-10 years.

Using 2003 Interim Egypt DHS data, Roushdy (2004) finds no significant relationship between these two when child nutrition is measured by Height-for-Age Z-score and women's autonomy is identified by mobility, acceptance of domestic violence, control over cash, and decision-making related to children. Another biological study finds a positive and significant relationship between maternal autonomy and Weight-for-Age Z-score in the sample of 820 pairs of mother and child aged 6 to 24 months in rural Karnataka, India (Sethuraman et al., 2006). In Shroff et al. (2011)'s study, 600 pairs of mothers and 3-5 months old infants from Andhra Pradesh, India are studied. Autonomy is defined by seven dimensions, including household decision making, mobility autonomy, acceptance of domestic violence, decision-making related to children, financial independence, actual mobility, and actual experience of domestic violence, and a Confirmatory Factor Analysis is conducted to attain indices for each dimension. The results suggest that when a mother has higher autonomy in household decision making, the infant is less likely to be underweight and wasted. Using similar measurement of autonomy, Arulampalam et al. (2016) find that high maternal autonomy could reduce the chance of child being stunted for children under 18 months in rural India, with a sample from the third round of India DHS.

Besides the challenges in the measurement of autonomy, the endogeneity issue of maternal autonomy also poses significant challenges to economists. To the best of our knowledge, only two studies attempt to deal with the endogeneity of women's empowerment. Lépine and Strobl (2013) suggest that bargaining power of a mother (measured by a woman's labour status, decision-making about herself and visits to her family, and ability to leave her house without permission) is positively associated with nutritional status of child in rural Senegal, using a mother's ethnicity relative to that of the community as an arguably exogenous IV. They exploit the fact that in the Senegalese context, Wolof women were more autonomous than Tukulor Fula women as they were Islamised later than the Fula, therefore a matrilineal political system had been sustained longer than the Fula. However, this relies on the assumption that such inter-ethnic marriage is not correlated with any unobserved factors that induce changes in child nutrition, such as beliefs of child upbringing. Another concern is that since political autonomy is a different dimension of autonomy from female autonomy at household level, which is their variable of interest, the strength of the link between these two (i.e. the relevance assumption for a valid IV) is questioned.

Using three rounds of India DHS (also known as National Family Health Survey) in 1992, 1998 and 2005, Imai et al. (2014) measure mother's empowerment by educational attainment relative to their husband, experience of domestic violence, and freedom to go to the market. They find a positive association between the above measures and child nutritional status. They use IV estimation for data collected in 2005, instrumenting empowerment by the proportional difference in age between father and mother and the village-level ratio of predicted average wage rates for women and men, yet the test of weak identification for the IV was not available. In our study, we adopt a new method without worry about finding a good IV to identify the causal effect of women's empowerment.

In the context of Ethiopia, existing studies use relatively small samples, and some find a positive association between women's empowerment and child nutrition. Using a small sample from rural Ethiopia, Deyessa et al. (2010) find the risk of child death increases when mothers with depression are exposed to domestic violence. In a case-control study, Garoma et al. (2012) show maternal intimate partner violence victimisation is strongly associated with under-five mortality in Western Ethiopia. Egata et al. (2014) reveal higher rates of wasting in couples who make an individual decision on a child's health in east rural Ethiopia. Tosheno et al. (2017) show that maternal decision-making power persists as a strong predictor of whether children are underweight aged 6-59 months in Southern Ethiopia. Demelash et al. (2015) find that those mothers suffering intimate partner violence during pregnancy were more likely to have a low-birth-weight child in Southeast Ethiopia. Fafchamps et al. (2009) find that improvements across several dimensions of female empowerment benefit the nutrition and education level of children using data in 15 villages in Ethiopia in the time period between 1993 and 1997. In addition to these case studies, comprehensive evidence, using up to date information, on the causal relationship between maternal autonomy and child nutrition in Ethiopia is still in demand.

3.3 Data

This study relies on the national representative Ethiopia Demographic and Health Survey (EDHS) (CSA and ICF, 2000-2016) to explore the causal effect of maternal autonomy on child nutrition. The sample for the EDHS is designed to provide estimates of population and health indicators for the country as a whole, the urban and rural areas separately, and each of the 11 regions in Ethiopia. There are four waves of EDHS in total, and in this study, the most recent three waves are studied, due to the lack of information on women's autonomy in the first wave. In the DHS surveys, women of reproductive age, between 15 and 49, are eligible for individual interview. Anthropometric measures to assess the nutritional status of children under age 5, and measures to proxy women's empowerment are available, as well as other child-level and sociodemographic characteristics. We restrict our sample to women who are currently in a monogamous marriage and residing with their husbands. This results in a sample consisting of 5119 pairs of mother-child.

3.3.1 Maternal autonomy: main explanatory variable

Dyson and Moore (1983) was one of the first studies to introduce the concept of female autonomy as the ability to control over her environment, and to obtain information, with her technical, social and psychological capability, which facilitates her in making decisions for her private concerns and intimate relationships. In the context of a household, it reflects women's capabilities in the realms of gathering information, making decisions and having control over choices that have an impact on her own life and her family (Carlson et al., 2015).

Meanwhile, the literature has documented that maternal autonomy is manifested over dimensions such as non-acceptance of domestic violence, household decision-making and control over sexual behaviours (Agarwala and Lynch, 2006; Sandberg and Rafail, 2013; Shroff et al., 2011), which are surveyed by the DHS. Recently, DHS data has been widely exploited in studies of women's autonomy by researchers who mostly use a summary score of the questions items to construct the index of autonomy. However, this simplification of measurement of autonomy has been critiqued by some as noted in the literature review. Carlson et al. (2015) argue that though different dimensions of autonomy are purported to reflect the same underlying concept, a woman having a high-level autonomy in one dimension would not necessarily induce a high level of autonomy in another dimension.

Richardson (2018) suggests using Confirmatory Factor Analysis (CFA), which acknowledges the multi-dimensional construct of autonomy and allows different loading on each item response to construct indices, rather than having a same weight assigned to each item response by simply aggregating them. In this study, we follow Richardson (2018) and Shroff et al. (2011) in choosing CFA model over Exploratory Factor Analysis (EFA), where the former is theory based and the latter is data based. The fact that key dimensions of autonomy have already been studied and recognised by previous studies enables us to rely on the CFA model, testing the applicability of the model in the sample.

3.3.1.1 Construction of latent autonomy variables: using Bayesian analysis based on a CFA model

Following Agarwala and Lynch (2006), Pratley and Sandberg (2018), Sandberg and Rafail (2013), and Shroff et al. (2011), we construct a second order measurement model in which the latent variable of maternal autonomy at the household level is operationalised according to the three factors of autonomy: autonomy in household decision-making; autonomy in non-acceptance of domestic violence; and autonomy in control over sexual behaviours.[5]

The index of non-acceptance of domestic violence is composed of five items based on whether the woman think domestic violence by husband is justified: (i) if she goes out without telling husband; (ii) if she neglects the children; (iii) if she refuses to have sex with husband; (iv) if she burns the food; and (v) if she argues with husband. Respondents are scored from 1 to 2: a woman is scored 1 if she accepts domestic violence; and she is scored 2 if she does not accept domestic violence.[6] The higher the score, the higher level of woman autonomy.

[5]A first order autonomy model with one latent variable maternal autonomy without any dimensions is also estimated, but the model fit is very low, rejecting the hypothesis that there is not a multi-dimensional construct of autonomy. So in the regression analysis, we will use the second order model as the CFA model suggests a good model fit.

[6]This is the same when coded starting from 0. The statistical model generates same result, using these two types of scoring systems.

TABLE 3.1: Percent of mothers answering autonomy questions in the sample

Autonomy in household decisions			
	Husband alone	Jointly with husband	Respondent alone
on women's health care	24.70	64.29	11.01
on large household purchases	31.96	61.08	6.96
on visits to women's family or relatives	18.49	69.74	11.77
on how to spend earnings by husband	26.29	70.06	3.65
Autonomy in non-acceptance of domestic violence			
		No	Yes
if wife goes out without telling husband		54.97	45.03
if wife neglects the children		60.50	39.50
if wife refuses to have sex with husband		44.57	55.43
if wife burns the food		55.30	44.70
if wife argues with husband		54.47	45.53
Autonomy in control over sex			
		No	Yes
woman can refuse sex		51.26	48.74
wife is justified to ask husband to use condom if he has		38.67	61.33
woman can ask partner to use a condom		64.89	35.11
Obs	5119		

Note: These percentages are weighted.

The index of household decision-making contains four factors: whether the woman can make a decision (i) on her health care; (ii) on large household purchases; (iii) on visits to women's family or relatives; and (iv) on how to spend earnings by the husband. Respondents are scored from 1 to 3: a woman is scored 1 if she does not make the decision, 2 if she and her husband jointly make the decision, and 3 if she makes the decision alone.[7]

The index of control over sex consists of three items: (i) whether the woman can refuse sex; (ii) whether the wife is justified to ask husband to use a condom if he has; and (iii) whether the woman can ask the partner to use a condom. Respondents are scored from 1 to 2: a woman is scored 1 if she does not have control over sex and 2 the otherwise.

The question items intent to measure these three indices we interest in are shown in Table 3.1, together with the summary statistics which suggests a low level of maternal autonomy in Ethiopia: there are only at least 44.57% of the sample accept domestic violence; less than 11.77% of the women can have the freedom to make decisions within household, and even with matters that are personal to women, such as their health care or visits to their family and relatives, it is the husband who is more likely to make decisions rather than women themselves; lastly, only 48.74% of wives can refuse sex, and 35.11% of them can take control of their reproductive decision by asking partner to use a condom when having sex.

Shown in Figure 3.1, we propose a second-order conventional CFA model, where maternal autonomy is the single construct with multiple dimensions, similar to Agarwala and Lynch (2006)'s model. The responses are on a 2 or 3 point Likert scale: a high score represents a high level of autonomy. To start

[7]As Mabsout (2011)'s finding indicates that the association is statistically significant with joint-decision making and women's health, an alternative of scoring the decision-making, i.e. making decisions jointly is scored as 3 while women making decisions alone as 2, and result of this new measurement of autonomy is available in the Appendix.

with, we use Mplus v.7.4 (Muthén and Muthén, 1998-2012) to carry out CFA and examine the model fit. The statistics for model fit are followed: RMSEA is 0.041; CFI is 0.990; TLI is 0.988. For RMSEA, a value under 0.05 suggests an excellent model fit; for CFI and TLI, values above 0.95 reaching to 1 indicate a good model fit (Brown, 2014). This suggests that the three-dimensional theoretical model fits well with the sample.[8]

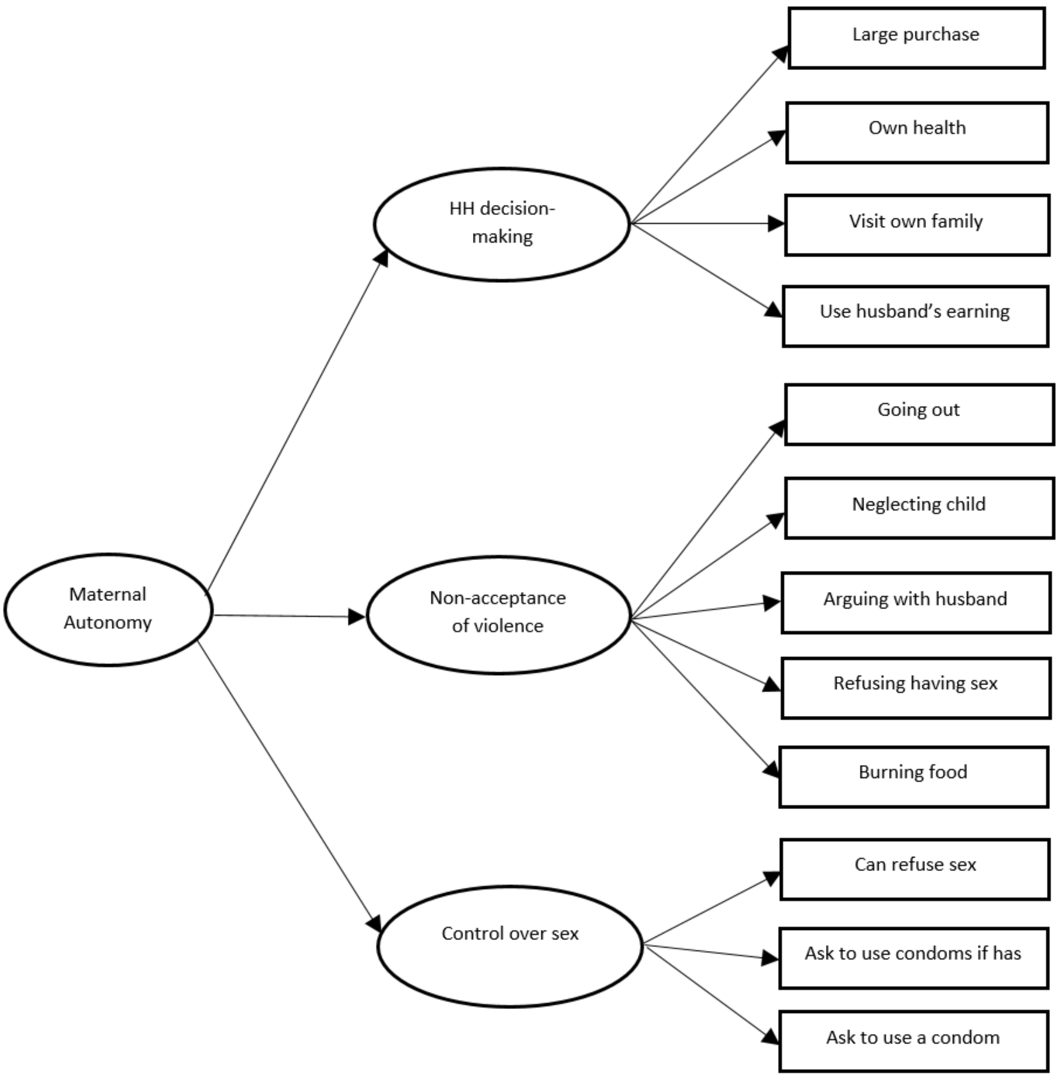
Since our interest is to obtain factor scores estimates of the latent variables to do further analysis, a Bayesian analysis based on the CFA model is utilised as the conventional CFA model in Mplus calculates factor scores using Item Responses Theory (IRT) for categorical items, which requires the number of items per factor variables to be at least 20. As the minimum number of items per factor variable is three in our model, a conventional CFA model cannot produce good factor score estimates, and thus we need a Bayesian plausible value approach based on the conventional CFA model to utilise our further analysis.

Another advantage of the Bayesian CFA is that it allows for relaxation of the priors conventional CFA model setting cross-loadings to be zero[9], which is a quite strict assumption. For example, in the conventional CFA setting, female autonomy in control over sex is assumed not to be correlated with autonomy in non-acceptance of domestic violence if refusing sex. However, the Bayesian CFA could set an informative and small-variance priors for cross-loadings, i.e. $\lambda \sim N(0, 0.01)$, such that the model relaxes the restriction to that the cross-loadings could be zero to 95% small cross-loading bounds of ± 0.20 (Muthén and Asparouhov, 2012). Applying the Bayes estimator and Gibbs algorithm, two independent Markov chain Monte Carlo with 100,000 iterations are used to describe the posterior distribution of the model parameters. Convergence is

[8]The diagram of conventional CFA model with factor loadings is shown in Appendix.

[9]Zero cross-loadings propose a none relationship between a factor and an item in the subset of another factor.

FIGURE 3.1: Conventional CFA model of autonomy



successfully reached within 100,000 iterations, as the potential scale reduction approaches to 1.00 and remains at such level.[10] In Figure D2, we report the calculated factor loadings using this model. We do find cross-relationship as we find a few significant cross-loadings. For example, it suggests a positive correlation between the latent variable of the dimension of control over sex and the observed variable of non-acceptance of domestic beating if refusing sex.

We assess the model fit by carrying out the Posterior Predictive Checking for the Bayesian models. A Posterior Predictive p-value (PPp) larger than 0.05 indicates good model fit, as PPp measures the proportion of the chi-square values of the replicated data that exceeds that of the observed data (Asparouhov and Muthén, 2010). In the model, the PPp equals 0.09, suggesting a good model fit.

Using the Bayesian CFA model, the factor scores are estimated with 100 iterations for each of the three dimensions of autonomy separately and for an overall composite combining them as an *autonomy* index. As shown in Table 3.2, similar with most of Agarwala and Lynch (2006)'s and Shroff et al. (2011)'s results, the three dimensions of autonomy are moderately correlated, with correlations of three factor scores estimates ranged from 0.30 to 0.37. This indicates that in the following regression analysis, different dimensions of autonomy could be treated as part of a single underlying construct, and there is a possibility of distinct contributions to child nutrition. Unlike perceptions of the legitimacy of violence being uncorrelated with other dimensions of autonomy in the context of India and Pakistan (Agarwala and Lynch, 2006), it is fairly correlated with other two dimensions in Ethiopia, suggesting women's opinion of domestic violence could plausibly reflect women's autonomy in

[10]The convergence process is shown by the potential scale reduction (Muthén and Asparouhov, 2012).

TABLE 3.2: Sample statistics for estimated factor scores using the Bayesian CFA model

	HH decisions	Control over sex	No domestic violence	Autonomy
HH decisions	1.000			
Control over sex	0.366*	1.000		
No domestic violence	0.363*	0.293*	1.000	
Autonomy	0.821*	0.708*	0.694*	1.000

Note: Pearson's correlation estimates are reported in this table. * p<0.05

Ethiopia. Meanwhile, the overall autonomy score is highly correlated with the three subdimensions, with a Pearson's r between 0.71 and 0.82. These autonomy measures are standardised in the following analysis.

3.3.2 Child nutritional status: outcome measures

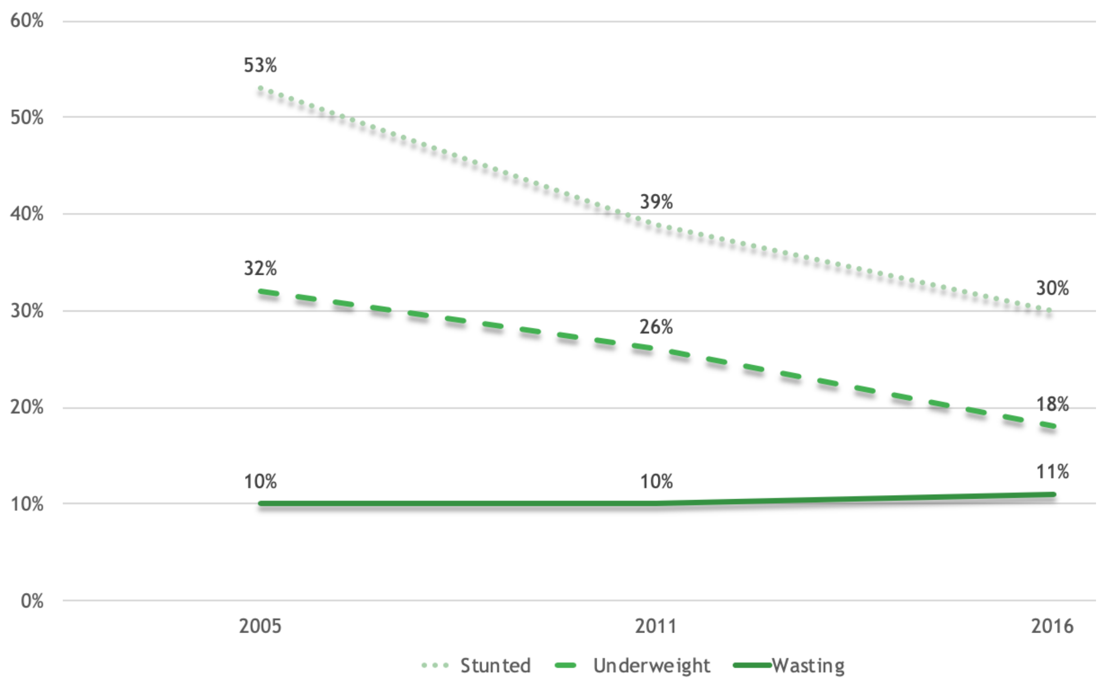
We assess nutritional status of child under age five using well established anthropometric indicators, including the height-for-age Z-score (HAZ) and weight-for-age Z-score (WAZ), calculated using the WHO 2006 growth standards (WHO, 2006).[11]

Two dummy variables are constructed as proxies for child undernutrition: stunting (HAZ less than -2) and underweight (WAZ less than -2), following WHO's definition of child mid- and long-term nutritional deprivation (WHO, 2010). Shown in Figure 3.2, child undernutrition has been reduced from 2005 to 2016 in Ethiopia, indicated by data from three rounds of DHS. Specifically, the rate of stunting decreases from 53 per cent to 30 per cent, and the rate of underweight drops from 32 per cent to 18 per cent, suggesting the risk of chronic malnutrition has fallen over the past 12 years, albeit at a slow rate.

[11]It is calculated using the Stata command *zscore06* developed by Leroy (2011).

This study explores whether elevated maternal autonomy has played a role in the reduction of malnutrition in Ethiopia.

FIGURE 3.2: Trends in nutritional status of children under age five, EDHS 2005-2016



3.3.3 Control variables

We consider child, maternal, parental, household and village characteristics as important confounding variables in the model. As we are using a novel method introduced in the next section, it allows us not to be parsimonious to include raw regressors representing these characteristics from the data as well as interactions terms and transformations of these regressors, with the aim of capturing and describing as much of the early endowment, investment and conditions of early childhood as possible.

According to previous studies in nutrition and women's empowerment, we select an extensive list of regressors z that are believed to have predictive power for either maternal autonomy or child nutrition as the potential control variables into our Lasso selection model. They are: (1) maternal traits described by maternal education, maternal height, maternal literacy level, maternal relationship to household head, maternal age at first birth, maternal age at first cohabitation, maternal age, currently working, and worked in last 12 months; (2) child characteristics captured by age in months, younger than 24 months, gender, first born, birth interval with the previous born less than 24 months, birth interval with the last born more than 24 months, the first time of breastfeeding after birth (within an hour, within a day, or more than a day), never been breastfed, whether it is a desired pregnancy, currently pregnant, year of birth, and month of birth; (3) paternal traits proxied by father's age, father's education level, difference of education years between mother and father, difference of age between mother and father; (4) household characteristics portrayed by region, religion, wealth, residence (rural or urban), average school years of mother and father, gender of household head, number of children under five, share of children under five, number of household, type of drinking water, type of toilet, summary of pieces of large assets, type of large assets, type of floor, type of roof, and type of cooking fuel; (5) village time-specific traits controlled by a village time-specific fixed-effects.

A full list of potential regressors z that are believed to have strong predictive power on both child nutrition and maternal autonomy are presented in the Appendix Table C1 and C2, while transformation and interactions between these characteristics are not shown.[12] To avoid the issue of "bad" controls (Angrist and Pischke, 2008), we do not include variables such as vaccination, current health status, size at birth, and place of delivery as they might be an outcome of maternal autonomy while associated with child nutrition.

[12]Child, maternal and parental characteristics are shown in Table C1, and household characteristics can be found in Table C2.

TABLE 3.3: Descriptive statistics for key variables

Variable	Mean	Std. Errors.
Maternal autonomy		
Control over sex	-0.177	0.025
Non-acceptance of violence	-0.083	0.024
Household decision-making	-0.095	0.027
Autonomy	-0.155	0.027
Outcome variables		
Underweight (dv)	0.254	0.010
Stunted (dv)	0.403	0.011
WAZ scores	-1.131	0.030
HAZ scores	-1.422	0.039
Child characteristics		
Child age in months	24.760	0.362
Younger than 24 months (dv)	0.529	0.010
Child male (dv)	0.504	0.011
First born (dv)	0.216	0.009
Birth interval less than 24 months (dv)	0.094	0.006
Birth interval between 24 and 48 months (dv)	0.423	0.010
Birth interval more than 48 months (dv)	0.267	0.010
Maternal characteristics		
Can't read (dv)	0.732	0.011
Child marriage (dv)	0.665	0.011
Age in years	28.844	0.160
Education: no (dv)	0.639	0.013
Paternal characteristics		
Age in years	36.225	0.223
Education: no (dv)	0.475	0.013
Difference in education years: mom-dad	-1.269	0.065
Difference in age in years: dad-mom	7.378	0.148
Household characteristics		
Rural (dv)	0.861	0.012
Religion: muslim (dv)	0.319	0.017
Wealth: bottom 20 percentile (dv)	0.201	0.012
Average school years of parents	2.527	0.092
Head: male (dv)	0.965	0.004
Head: age in years	37.153	0.234
Share of kids under five in family	0.269	0.003
Size of household	5.689	0.050
Drinking uncleanwater: (dv)	0.425	0.015
Having no toilet facility (dv)	0.451	0.015
Floor is made of sand or dung (dv)	0.885	0.008
Roof is none, or made of thatch or mud (dv)	0.395	0.014
Cooking fuel: wood, dung, grass, or crop (dv)	0.918	0.007
Total pieces of large assets	0.678	0.027
No. of large asset (dv)	0.571	0.013
No. of observations	5119	

Note: "*Dv*" is denoted for dummy variable. The means and standard errors of variables are estimated using the weights from EDHS.

Table 3.3 presents statistics of some key characteristics of the child, parents, and household in the sample. Half of the children in the sample are male, and half of them are aged 0-23 months. More than half of mothers are married younger than 18, indicating the prevalence of child marriage, while on average their partners are seven years older than them. There are 64% of mothers without education and the mean of school years of parents is only 2.5 years. The majority of the households are from rural areas and headed by a male. The raw material of floor and roof, as well as the unclean source of drinking water, rare use of processed cooking fuel, and the absence of toilet facility and large assets, indicates the poor living standard of the sample.

3.4 Econometric model

Knowing that our variable of interest autonomy index is not randomly assigned, a classic method to study the causal effect of such on observational data relies on the assumption that the interest variable can be regarded as randomly assigned conditional on a sufficient set of factors. Conventionally, an *ad hoc* sensitivity analysis is usually utilised to show that the size and sign of parameter of interest variable do not change much when controlling different sets of variables. However, though economic intuition gained from the Early Childhood Development literature (Currie and Almond, 2011; Almond et al., 2018) would suggest us to control for variables that explain variations in nutrition (such as household composition, living conditions, household assets, quality of household construction, wash and sanitation practices, child caring practices, and village services), there is no clear guidance about exactly which variables should enter the human formation model, and how to construct confounding variables so that the early investments and conditions contributing to child nutrition are best described.

As a result, this raises concern that variables are selected consciously or unconsciously by researchers in order to produce good performing estimates of the parameter of interest (i.e. an estimate with statistical significance and the expected sign: Brodeur et al. (2018), Gelman and Loken (2013), and Vivalt (2018)), which is also an unsolved problem in the development literature. To eliminate the concern of "*p*-hacking" (Gelman and Loken, 2013), a novel estimation and uniformly valid inference method, called the "post-double-selection Lasso" (PDS-LASSO), has been developed by Belloni et al. (2014), allowing for including available raw regressors, interactions and transformations of regressors into the model, without worrying about the overfitting problem.[13] We exploit this method to study the causal relationship between autonomy and child nutrition in this setting.

The PDS-LASSO firstly selects controls via double feasible Lasso (least absolute shrinkage and selection operator) methods: we use one Lasso to select controls that have a strong predictive power for the interest variable (autonomy), and use another Lasso to select controls that are strongly related to the outcome variable (child nutrition). [14] It then estimates the effect of autonomy on child nutrition using OLS, including the selected controls from the double

[13]An overfitting problem occurs when a statistical model contains too many parameters to be justified by the data.

[14]The variable selection steps adopt the method of least absolute shrinkage and selection operator (Lasso) Tibshirani (1996). $\hat{\boldsymbol{\beta}}_{\text{lasso}}(\lambda) = \arg\min \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}\right)^2 + \frac{\lambda}{n}\sum_{j=1}^{p}\psi_j\left|\beta_j\right|$, where the tuning parameter λ controls the overall penalty level and ψ_j are predictor-specific penalty loadings. The PDS-LASSO uses a theory-driven penalty in the 'rigorous' penalisation, developed by Belloni et al. (2012). This method firstly estimates the penalty level λ for LASSO, and then obtain the LASSO coefficients using feasible algorithms. The optimal λ can also be obtained under heteroskedastic and cluster cases (Belloni et al., 2016). Therefore, we use the default λ in the cluster-lasso cases here, where penalty loadings differ across regressors. More details of the method are discussed in Belloni et al. (2012, 2014, 2016), and we only briefly describe it here. $\lambda = \lambda_0 \times rmse$, where rmse is an estimate of the standard deviation of the error variance, and $\lambda_0 = 2c \times sqrt(N) \times invnormal(1 - (\gamma/log(N_{clust}))/(2p))$, where p is the number of penalised regressors, c and γ are constants with default values of 1.1 and 0.1, N_{clust} is the number of clusters which is 64 in my analysis, and 'invnormal' means an inverse cumulative standard normal distribution. The penalty loadings for the cluster-robust case are, $\psi_j = sqrt[avg(u_i^2)]/sqrt[avg(e^2)]$, where u_i is the sum of $x_{ij} \times e_{ij}$ over the j members of cluster i, x is a demeaned regressor, and e is the residual. A sampling weight in the DHS data is also used in the LASSO estimation.

Lasso. The steps are shown as follows,

Step 1: Use the Lasso to estimate

$$H_{impt} = Z_{impt}\Pi + \gamma_p + \zeta_{impt} \tag{3.1}$$

where the child nutritional outcomes H_{impt} for child i of mother m in primary sampling unit (*psu*) p at time t is the dependent variable; Z_{impt} represents potential set of controls on which we need to condition, including the child, maternal, parental, and household characteristics; γ_p is the unpenalised fixed-effects to control for confounding factors at *psu*-wave level; and ζ_{impt} is the disturbance.[15] Note that our interest variable autonomy is not in this model. Denote the set of Lasso-selected controls from Z_{impt} by W_1.

Step 2: Use the Lasso to estimate

$$A_{mpt} = Z_{impt}\Psi + \gamma_p + \eta_{impt} \tag{3.2}$$

where the maternal autonomy A_{mpt} for mother m in *psu* p at time t is the independent variable; Z_{impt} is the same as above; γ_p is the unpenalised fixed-effects at *psu*-wave level; and η_{impt} is the disturbance. Denote the set of Lasso-selected controls by W_2.

Step 3 (main model): Use OLS to estimate the following fixed effects model

$$H_{impt} = \beta A_{mpt} + X_{impt}\Theta + W_{impt}\Lambda + \gamma_p + \epsilon_{impt} \tag{3.3}$$

where H_{impt} represents child nutrition, and A_{mpt} denotes for maternal autonomy; X_{impt} are a set of unpenalised child-level controls that we believe are essential in the model and hence will not be excluded, i.e. gender, birth year dummies, birth month dummies, rural or urban, child's age in months and

[15]In DHS, an enumeration area (EA) is a geographic area covering on average 181 households, and a cluster is either an EA or a segment of an EA. Each of the clusters or *psu*s was displaced within 2 kilometres from the actual location for urban points and within 10 kilometres for rural points. In EDHS, the *psu*s are not the same across rounds. There are 1522 *psu*s in our sample. Thus the *psu* fixed-effects model controls for the unobserved time-specific village traits.

square of child's age in months, and birth interval dummies[16]; W_{impt} are the set of union of the Lasso-selected controls from Step 1 and Step 2 that are related to child nutrition and maternal autonomy, i.e. $W_{impt} = W_1 \cup W_2$; γ_p is the unpenalised fixed-effects at *psu*-wave level, capturing two strings of confounding variables: (1) local gendered institutions which affect women's autonomy, such as the asymmetric social norms and belief (Fafchamps et al., 2009; Goetz, 1997; Staveren and Ode bode, 2007); (2) village specific characteristics that are related to child nutrition, such as cleanness of living conditions in villages, availability of health service and traditional health practice; and ϵ_{impt} is the disturbance. The inference is conducted based on clustered standard errors at region-wave-rural level. There are 64 clusters and 1522 *psu*s in our sample.

Our interest is to find out the causal effect of maternal autonomy on child nutrition, which is potentially revealed by the coefficient β. However, the possibility that unobserved heterogeneity, such as household socio-economic status, father's gender norms, and household environment, might raise a concern of bias in the estimates. We make the following assumptions in order to validate an unbiased causal effect depicted by coefficient β when we apply LASSO to a village-wave fixed effects model. Firstly, the village-wave fixed effects successfully capture any unobserved heterogeneity in father's gender norms (as it is formed mainly by the local culture and environment), village gendered institutions, village socio-economic characteristics, and time trend. Secondly, comparing to the manual selection of control variables that portray household characteristics, the LASSO-selected variables best describe household environment and parental socio-economic status, without worrying missing any significant variables in explaining outcomes. The innovative PDS-LASSO is

[16]We show results of models in which these variables are not unpenalised as robustness check in Table C6. Specifically, we do not select these variables manually and do not put these into X_{impt}, but into Z_{impt}, allowing LASSO to freely select indicators that are related to the interest variable and dependent variable.

designed to deal with the omitted variable bias problem in estimating causal effects; it allows us to use a partially linear model and employing potentially high-dimensional controls, but at the same time avoiding the risk of overfitting and imperfect selection of abundant control variables (Belloni et al., 2014).

3.5 Results

3.5.1 Preliminary results

Before presenting the main results, we show preliminary results in Table 3.4, where a series of controls are selected manually and included into the model in sequence using the conventional approach. We report here the results from specifications where maternal autonomy is measured by household decision-making as an example. In Panel A, the results for being underweight as the dependent variable are shown; in Panel B, the ones for being stunted as the outcome are displayed. All models in column 1 to 5 include a *psu*-wave fixed-effects.

In general, the coefficients are negative, in line with the hypothesis. As shown in column 2, when unpenalised variables in PDS-LASSO model (i.e. birth year, birth month, rural, age in months, gender, first born, and birth interval) are added in the models, the coefficients are negative and significant at 10% level in both panels. In column 3 to 5, we report results when some variables at child, mother, father and household level[17] are manually selected and

[17]The controls are breastfed; mother's literacy level, mother's relationship to the head of household, mother's age, mother's education level, mother's height, mother's age when having first born baby, and mother's age when first cohabitation; father's age, father's education level, different education years between father and mother, and age difference between father and mother divided by average age; wealth index, share of under-five children in the household, household size, average years of education, gender of household head, using unclean water, no toilet, no large asset, poor material of floor, and poor material of roof.

included in the model sequentially. In column 3, the coefficients are still significant at 10% level and the magnitudes remain similar to the ones in column 2, when breastfeeding is being controlled for. However, in column 4, it appears that as more controls are added, the standard errors and sizes of coefficients are not stable, and the coefficients are no longer not significant at 10%. In column 5, three plausible 'good' controls (i.e. birth interval, mother's age when having first born baby, and mother's age when first cohabitation) are dropped as a robustness check, in case that they take away the 'effect' from the outcome variables.[18] The results are similar to the ones in column 4.

The preliminary results suggest that, with a large choice of variables at hand, researchers could report desired results by choosing specific controls that produce statistically significant coefficients, ignoring the unstable performance of estimates. This convinces us that imperfect variable selection would be problematic for establishing robust evidence. It therefore motivates us to use the PDS-LASSO approach to test whether the results persist without manual selection of control variables, since there is no clear guidance about what variables we should control for besides the unpenalised ones, i.e. those that are clearly motivated for inclusion through the economic model. The production of nutrition relies on individual traits, such as age (Akombi et al., 2017), birth year and month (Branca et al., 1993), gender (Svedberg, 1990), birth spacing (Lindstrom and Berhanu, 2000), and place of residence (urban or rural) (Akombi et al., 2017), and we take them into our models as the unpenalised control variables.

[18]There is a concern of whether birth interval, mother's age when having first born baby, and mother's age when first cohabitation are outcomes of maternal autonomy and correlated with the nutritional outcome variable, which would take away the 'effect' from outcomes. So that we drop these three variables to check the changes in results.

TABLE 3.4: FE regression models: preliminary results

Panel A: Dependent variable	Being underweight (dv) (Mean=0.254)				
	(1)	(2)	(3)	(4)	(5)
Household decision-making	-0.020**	-0.018*	-0.018*	-0.010	-0.011
	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)
Psu-wave FE	Y	Y	Y	Y	Y
Unpenalised variables in PDS-LASSO		Y	Y	Y	Y
Breastfed controls			Y	Y	Y
Parent and HH controls				Y	Y
Drop some plausible 'good' controls					Y
Panel B: Dependent variable	Being stunted (dv) (Mean=0.403)				
	(1)	(2)	(3)	(4)	(5)
Household decision-making	-0.016	-0.015*	-0.016*	-0.007	-0.008
	(0.011)	(0.009)	(0.009)	(0.010)	(0.011)
Psu-wave FE	Y	Y	Y	Y	Y
Unpenalised variables in PDS-LASSO		Y	Y	Y	Y
Breastfed controls			Y	Y	Y
Parent and HH controls				Y	Y
Drop some plausible 'good' controls					Y
Observations	5119	5119	5119	5119	5119

Note: Region-wave-rural clustered standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All models apply a *psu*-wave fixed-effects estimation. The unpenalised variables are birth year, birth month, rural, child's age in months, gender, first born, and birth interval. The breastfed controls are never been breastfed, breastfed within an hour after birth, breastfed within a day after birth, and breastfed more than a day after birth. The parent controls include literacy level, relationship to the head of household, age, education level, mother's height, age when having first born baby, age when first cohabitation, different education years between father and mother, and age difference between father and mother divided by average age. The household-level controls include wealth index, composition of household, average years of education, gender of household head, using unclean water, no toilet, no large asset, and material of roof and wall. The model in column 5 does not include some plausible 'good' controls, i.e. birth interval, and mother's age when having first born baby and when first cohabitation. These coefficients of regressions and the mean of outcome variables are estimated using the weights from EDHS.

3.5.2 Main results

We use Stata 14 and the PDS-LASSO package written by Ahrens et al. (2018) and Ahrens et al. (2019) to run the regressions. Since the models adopt *psu*-wave fixed effects, any time trend, regional and village effects are purged from the error component. Each specification is estimated using the PDS-LASSO techniques, and thus the variable selection is conducted separately for each specification. For example, as the outcome variable and the interest variable are changed in each specification, the series of selected variables may vary. Details of what variables are selected as controls by the LASSO can be found in Appendix Table C3 and C4.

In Table 3.5, results from regressions employing being underweight (dummy) and WAZ scores as outcome variables are reported in Panel A, while the ones using being stunted (dummy) and HAZ scores as dependent variables in Panel B. Before running the regressions, the correlation between potential controls and interest variables are tested using Pearson's correlation test. The highest correlation is no more than 0.43. In column 1 and 6, the impact of the composite score of autonomy is studied; in column 2 to 4 and column 7 to 9, each index of three autonomy dimensions, i.e. household decision-making, non-acceptance of domestic violence and control over sex, is separately regressed on nutrition status; and in column 5 and 10, all indices of three dimensions are jointly included in the specification.

As expected, the results indicate that measured by various indices, autonomy impacts negatively on poor nutrition indicators, i.e. being underweight and stunted, but positively on the anthropometric measures. Specifically, shown in Panel A column 1, 3 and 5, the estimates from specifications depicting inadequate nutrition by underweight are statistically significant. In column 1 of Panel A, the result suggests that a standard deviation increase in overall maternal autonomy will cause a 2% decrease in the likelihood of the child being underweight, as the estimate is -0.020 with a 90% confidence interval of [-0.038, -0.002]. Despite of a rather wide confidence interval, this effect size is important and comparably large, since as shown in Figure 3.2, the incident of overall underweight only drops by 1.17% per year from 2005 to 2016, and there are still 18% children at risk in 2016. The size is relatively close to Smith et al. (2003), which reveals that in Sub Saharan Africa (not including Ethiopia in their study) if women share the same status as men, child malnutrition would reduce by almost 3%.

Next we dig down to see which aspect of autonomy contributes most to this effect. In column 2 of Panel A, the estimate of autonomy in household decision-making is negative but statistically insignificant, with a size of -0.011 and a 90% confidence interval of [-0.026, 0.004]. This confirms what we find in preliminary results that the coefficient for autonomy in household decision-making is not statistically stable when control variables vary. The effect 'disappears' when using the new methodology. We do not find similar results as Tosheno et al. (2017)'s, which indicate that maternal decision-making ability is a strong predictor of child being underweight aged 6-59 months, using a community-based cross-sectional dataset in Ethiopia.

However, in column 3 of Panel A, the estimate is -0.020 with a 90% confidence interval of [-0.033, -0.007], suggesting that when there is one standard deviation increase in maternal non-acceptance of domestic violence, the probability of the child being underweight decreases by 2% at 95% confidence level. This coefficient remains relatively stable when the model consists of all three autonomy dimensions, as the estimate is -0.018 with a 90% confidence interval of [-0.031, -0.005], reported in column 5. The findings are consistent with the results from Basu and Koolwal (2005), which find that when mothers do not think domestic violence is justified, the youngest child is significantly less

TABLE 3.5: PDS-LASSO regression models: main results

Panel A:	Being underweight (dv) (Mean=0.254)					WAZ (Mean=-1.131)				
Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Autonomy	-0.020*					0.009				
	(0.011)					(0.038)				
Household decision-making		-0.011			-0.003		-0.005			-0.015
		(0.009)			(0.008)		(0.029)			(0.028)
Non-acceptance of violence			-0.020**		-0.018**			0.033		0.034
			(0.008)		(0.008)			(0.040)		(0.040)
Control over sex				-0.013	-0.009				0.003	-0.004
				(0.008)	(0.008)				(0.020)	(0.019)
Panel B:	Being stunted (dv) (Mean=0.403)					HAZ (Mean=-1.422)				
Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Autonomy	-0.003					0.037				
	(0.011)					(0.033)				
Household decision-making		-0.007			-0.007		0.029			0.022
		(0.010)			(0.009)		(0.038)			(0.043)
Non-acceptance of violence			-0.006		-0.004			0.017		0.005
			(0.008)		(0.007)			(0.026)		(0.031)
Control over sex				-0.002	0.001				0.042	0.034
				(0.009)	(0.008)				(0.030)	(0.030)
N					5119					

Note: Region-wave-rural clustered standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All models apply a *psu*-wave fixed-effects to the PDS-LASSO model. The unpenalised variables are birth year, birth month, rural, child's age in months, gender, first born, and birth interval. These coefficients of regressions and the mean of outcome variables are estimated using the weights from EDHS.

likely to be anaemic. The results are also in line with the idea suggested by Fafchamps et al. (2009) that, in the context of Ethiopia, the predisposition towards domestic violence may exhibit a significant role in intrahousehold distribution of food, as they find a negative association between domestic violence and wife's BMI using household fixed-effect.[19]

Combined with the Self Determination Theory discussed earlier, we believe that women's acceptance of domestic violence reveals women's motivations behind child caring behaviours to be extrinsic, hence disabling willing practices for child care. This autonomy measure is depicted by questions, such as whether women accept being beaten if they neglect children and whether women accept being beaten if they burn the food. Positive response implicitly confirms that women experience fear when they are performing these events, and quality of such behaviours might be driven by fears of being punished, rather than an intrinsic enjoyment and desire of doing so. Consequently, high acceptance of domestic violence reflects an extrinsic motivation in taking care of children, which decreases the quality and desire of child-caring in the long run.

In column 4 of Panel A, it shows that the estimate of autonomy in control over sex equals to -0.013, with a 90% confidence interval of [-0.026, 0]. As for results of specifications in column 6 to 10, the standard errors are large, failing to reject the null hypothesis that the coefficients are zero.

As for results in Panel B, the estimates are with expected sign (i.e. negative for outcomes proxied by stunted and positive for outcomes proxied by HAZ scores), suggesting malnutrition is negatively associated with autonomy, yet not statistically significant. The standard errors are quite large, resulting in wide confidence intervals, except the one in column 9. In the specification

[19]The predisposition towards domestic violence in Fafchamps et al. (2009)'s paper is calculated using information from two variables: (i) if the respondent has ever seen their mother being beaten by their father; and (ii) if the respondent has ever involved in a physical fight.

regressing HAZ on autonomy of control over sex, the estimate is 0.042 with a 90% confidence interval of [-0.008, 0.092]. The null results of regressions using being stunted/HAZ might be due to the reason that these indicators proxy the long-term deprivation in nutrition, while contemporary maternal autonomy might not be able to impact on the accumulative stock of nutritional status, but could only reflect on the mid-term nutritional status, i.e. being underweight. These findings echo with those studies only finding that autonomy is correlated with underweight/WAZ, but not with stunted/HAZ (Roushdy, 2004; Sethuraman et al., 2006; Shroff et al., 2011). The effect is also reasonably sizeable, since impacts of some early childhood interventions, such as a US $60 million nutrition programme in Bangladesh (Hossain et al., 2005), are not found in reducing the prevalence of underweight.

In summary, these findings indicate that in general, maternal autonomy can benefit child nutrition conditional on child traits, parental characteristics, household characteristics, time trend, and village environment in Ethiopia. The evidence reconciles with Fafchamps et al. (2009)'s conclusion that empowered mothers might allocate household resources to invest in both nutrition and education of children in rural Ethiopia. It sheds light on improving mother's autonomy as a potentially strong pathway to improve mid-term child nutritional status and hence later life outcomes, such as educational outcomes, health, and income, in Ethiopia. Recalling the low maternal autonomy status shown in Table 3.1, we think that there is sufficient room in Ethiopia to enhance women's empowerment, which is linked to intergenerational transmission through child nutrition.

3.5.3 Heterogeneous results

The results above show an aggregate effect of autonomy on child nutrition. In the following, this study expands the analysis from Table 3.5 Panel A column 1 to 5 to explore the heterogeneous effect of autonomy concerning variation in two exogenous child indicators, i.e. child's age and child's gender.

Prior to adding interaction term into regression, t tests are carried out to see whether there is a systematic difference in autonomy between groups of mothers with younger children (0-23 months) and with older children (24-59 months), and between groups of mothers with daughters and sons. The t statistic from the previous test equals to -1.40, suggesting the difference in means of autonomy between groups with younger and older children is not statistically significantly different from zero. With a t statistic of -1.21 from the latter test, it indicates the mean of autonomy of mothers paired with sons is not statistically higher than the ones with daughters, although it is expected that women having a son might enjoy a higher level of autonomy in a son-preference society. These results provide evidence showing that maternal autonomy is independent with the age and gender of the child.

Shown in Table 3.6, specifications added interaction terms of autonomy with a young child are shown in Panel A. In column 1, the magnitude of coefficient of the overall autonomy score increases from 0.020 to 0.036, compared to the result from regression without interaction term (column 1 of Panel A in Table 3.5), suggesting that autonomy might have a larger impact on those who are older than 24 months. However, this result is not verified as the interaction term is not significant albeit positive.

Nonetheless, heterogeneous effects are found when using measures proxied by subdimensions. Different from the results in Table 3.5, the estimate of household decision-making, as shown in column 2, is statistically significant

TABLE 3.6: PDS-LASSO regression models: heterogeneous results

Panel A: young child differential	Being underweight (dv) (Mean=0.254)			
(=1 if younger than 24 months)	(1)	(2)	(3)	(4)
Autonomy	-0.036**			
	(0.017)			
Autonomy × young child	0.029			
	(0.021)			
Household decision-making		-0.024**		
		(0.012)		
Household decision-making × young child		0.026*		
		(0.015)		
Non-acceptance of violence			-0.040**	
			(0.017)	
Non-acceptance of violence × young child			0.040*	
			(0.022)	
Control over sex				-0.014
				(0.015)
Control over sex × young child				0.001
				(0.020)
Panel B: male child differential	Being underweight (dv) (Mean=0.254)			
(=1 if child is male)	(1)	(2)	(3)	(4)
Autonomy	-0.035**			
	(0.018)			
Autonomy × male child	0.029			
	(0.028)			
Household decision-making		-0.015		
		(0.013)		
Household decision-making × male child		0.010		
		(0.019)		
Non-acceptance of violence			-0.037**	
			(0.016)	
Non-acceptance of violence × male child			0.032	
			(0.033)	
Control over sex				-0.022*
				(0.014)
Control over sex × male child				0.020
				(0.024)
N			5119	

Note: Region-wave-rural clustered standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All models apply a *psu*-wave fixed-effects to the PDS-LASSO model. Each autonomy index is interacted with dummy variable young child ($= 1$, if child is younger than 24 months) in Panel A and with dummy variable male child ($= 1$, if child is male) in Panel B. The unpenalised variables are birth year, birth month, rural, child's age in month, gender, first born, and birth interval. These coefficients of regressions and the mean of outcome variables are estimated using the weights from EDHS.

after inclusion of the interaction, with a size of -0.024 and a standard error of 0.012 ($z = 2.00$), and the estimate of the interaction is significant and positive with a size of -0.026 and a standard error of 0.015 ($z = 1.73$). It suggests that when maternal autonomy in decision-making increases by one standard deviation, of whose child is older than 24 months, he/she is 2.4% less likely to be underweight. This study indicates that higher maternal autonomy in household decision-making reduces the occurrence of child being underweight among older group of children aged 2-5 years, yet not for children younger than two. It is consistent with Brunson et al. (2009)'s results using data from northern Kenya, which only find significant and positive relationship between autonomy and WHZ in the group aged 3-10 years, but not in the group of younger children.

The differential effect might be a result of the breastfeeding practice for children under 2, when primary nutrition input is breast milk, which is provided solely by mothers without the need for negotiating with partners or displaying autonomy. However, when children are weaned, it is essential for mothers to provide adequate type and amount of food to maintain children's growth. Mothers with less ability in decision-making in the family would have a lower power to distribute sufficient food to children, which renders nutritional deprivation for children during the critical developmental period. However, Shroff et al. (2011) find that mothers with autonomy in decision-making have infants less underweight in Andhra Pradesh, India, using a sample of children 3-5 months.

In column 3, when women's perception of domestic violence is studied, a similar result is found, suggesting one standard deviation increase in autonomy of non-acceptance of domestic violence would reduce the probability of a child being underweight by 4% among older group, but a zero correlation in the younger group. In column 4, no significant results are found, using autonomy in control over sex.

Shown in Panel B, results from specifications including interactions with the gender of the child fail to indicate that there is heterogeneous effect between mothers with daughters and sons, although the coefficients of autonomy itself increase and significant, compared with the ones without inclusion of the interaction (see column 1, 3 and 4 in Panel B). Although the sign of the interaction term is positive, it is not statistically significant. This result is not able to support previous evidence that Ethiopian mothers incline to invest in sons, who are deemed to take care of mothers in old age (Quisumbing and Maluccio, 2003).

Finally, to provide indicative evidence on the mechanism of this heterogeneous effect by the age of the child, we use the sample of older group and regress decision-making autonomy on the summary of types of staple food that child ate, which might proxy for the food security of child.[20] Shown in Table C5, the results indicate a positive relationship between maternal autonomy in decision-making and adequate dietary intake of the child, conditional on a series of the child, mother and household controls, in a sample of 320 mother-child pairs. It suggests that a mother who has a higher ability to make household decisions is more likely to provide adequate food to the child.

3.5.4 Robustness checks

A battery of robustness checks is conducted to test the consistency of the results. Firstly, in the first two variable selection steps, we add those unpenalised variables in the main models (last step using OLS) back to selection, i.e. birth year, birth month, rural, child's age in months, gender, first born and

[20] Specific diet information is not available in the EDHS for children older than 24 months. So we could only use the types of staple food for children as a proxy for food security.

birth interval. Although after the double-selection LASSO estimation, a different set of variables is provided to be controlled for in the OLS estimation, the results are consistent with the previous main results, shown in Table C6.[21]

Reckoning the existence of pastoral areas in Ethiopia, namely Afar and Somali, where familial structure and behaviour might deviate largely from main areas in the country, estimations studying the heterogeneity of the effect of autonomy are re-do using a restricted sample which excludes these two regions. The results, shown in Table C7, replicate the ones presented in the last section, suggesting that the main effect exists mainly among the older children.

As Mabsout (2011) suggests that women who make decisions jointly might have better well-being in Ethiopia, autonomy is calculated in the alternative way. Women making decisions jointly is assigned with the highest score, and new results using this new measure are reported in Table C8. The new findings are not dissimilar to main results.

In Table C9, results using a restricted sample of women with only one child under five are showed. They displays the same direction of causal relationship between women's status and child malnutrition.

3.6 Conclusion

This paper investigates whether women's autonomy impacts on child nutritional status using nationally representative data on Ethiopia. A PDS-LASSO method combined with village-wave fixed effects is employed to account partially for the potential endogeneity raised by omitted variables, enabling us to make use of the raw regressors from the survey data to proxy for as many the inputs of human capital development in the early childhood as possible,

[21]For example, the estimate of non-acceptance of violence is -0.020 in Table 3.5, while -0.017 in Table C6.

without worrying about either unintended researcher bias or having a sparse model. We find that a woman with high autonomy can improve the welfare of her child, in terms of better nutritional outcomes. One standard deviation increase in maternal autonomy can reduce the likelihood of child being underweight by 2%. Moreover, this effect is strongest for the woman with a child aged 24-59 months. Since breastfeeding practice in Ethiopia is prevalent and natural for a child under two, an infant's nutritional intake is not deprived by the lack of women's autonomy. However, adequate child-caring and feeding practice for a weaned child might require more of mother's willing plans, intended behaviours, and the ability to act upon the plans. These findings, serving as evidence of a potential link of enhancing child nutritional outcome in Ethiopia, raise the awareness of improving mother's autonomy, as it does benefit not only the welfare of women but also one of children.

The study complements to Early Childhood Development literature as it calls attention to the intangible psychological input of nutrition production. It indicates that maternal autonomy in child caring and mother's intrinsic motivation for nurturing a child might be as important as sufficient resources and adequate nutritional intakes in child early life development.

Further analyses on how specifically autonomy is related to health-seeking and child-caring behaviour, and which dimension of autonomy is of high importance to nutritional outcomes could be explored. In our analysis, we find that autonomy in non-acceptance of domestic violence reduces the probability of a child being underweight. Women's acceptance of domestic violence might not only reflect women's belief in their unequal right to men, but also reveal that women's motivation behind their behaviours is extrinsic, which therefore disables willing behaviours for child care. For instance, one question asks whether women accept being beaten if they neglect children. A positive answer would indicate that women might exhibit child caring behaviours out of the fear of being punished if they do not. As a result, it would shift mother's motivation in taking care of children from an intrinsic, self-interest one towards an extrinsic, self-protective one, and hence might decrease the quality and desire of child-caring. This result has policy implications highlighting the importance of improving female autonomy, which stimulates intrinsic motivation for child caring, and hence provides insight into child nutritional state in Ethiopia, stressing that not only the quantity of nutritional inputs matters but also the psychological state of child carer matters.

Bibliography

Agarwala, Rina and Scott M Lynch (2006). “Refining the measurement of women’s autonomy: An international application of a multi-dimensional construct”. *Social Forces* 84.4, pp. 2077–2098.

Ahrens, A., C B Hansen, and M E Schaffer (2019). *LASSOPACK: Model selection and prediction with regularized regression in Stata*. Tech. rep. arXiv:1901.05397.

Ahrens, Achim, Christian B Hansen, and Mark E Schaffer (2018). *LASSOPACK: Stata module for lasso, square-root lasso, elastic net, ridge, adaptive lasso estimation and cross-validation*. `http://ideas.repec.org/c/boc/bocode/s458458.html`. Accessed: 2019-01-25.

Akombi, Blessing, Kingsley Agho, John Hall, Nidhi Wali, Andre Renzaho, and Dafna Merom (2017). “Stunting, wasting and underweight in sub-Saharan Africa: A systematic review”. *International Journal of Environmental Research and Public Health* 14.8, p. 863.

Allendorf, Keera (2012). “Women’s agency and the quality of family relationships in India”. *Population Research and Policy Review* 31.2, pp. 187–206.

Almond, Douglas, Janet Currie, and Valentina Duque (2018). “Childhood circumstances and adult outcomes: Act II”. *Journal of Economic Literature* 56.4, pp. 1360–1446.

Angrist, Joshua D and Jörn-Steffen Pischke (2008). *Mostly harmless econometrics: An empiricist’s companion*. Princeton University Press.

Arulampalam, Wiji, Anjor Bhaskar, and Nisha Srivastava (2016). *Does greater autonomy among women provide the key to better child nutrition?* IZA Discussion Papers 9781. Institute for the Study of Labor (IZA).

Asparouhov, Tihomir and Bengt Muthén (2010). “Bayesian analysis using Mplus: Technical implementation”.

Basu, Alaka Malwade and Gayatri Brij Koolwal (2005). "Two concepts of female empowerment: Some leads from DHS data on women's status and reproductive health". *A focus on gender–Collected papers on gender using DHS data, ORC Macro, Calverton MD*, pp. 15–33.

Belloni, Alexandre, Daniel Chen, Victor Chernozhukov, and Christian Hansen (2012). "Sparse models and methods for optimal instruments with an application to eminent domain". *Econometrica* 80.6, pp. 2369–2429.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen (2014). "Inference on treatment effects after selection among high-dimensional controls". *The Review of Economic Studies* 81.2, pp. 608–650.

Belloni, Alexandre, Victor Chernozhukov, Christian Hansen, and Damian Kozbur (2016). "Inference in high-dimensional panel models with an application to gun control". *Journal of Business & Economic Statistics* 34.4, pp. 590–605.

Branca, F, G Pastore, Tsegaye Demissie, and A Ferro-Luzzi (1993). "The nutritional impact of seasonality in children and adults of rural Ethiopia". *European Journal of Clinical Nutrition* 47.12, pp. 840–850.

Brodeur, Abel, Nikolai Cook, and Anthony Heyes (2018). *Methods Matter: P-Hacking and Causal Inference in Economics and Finance*. IZA Discussion Papers 11796. Institute for the Study of Labor (IZA).

Brown, Timothy A (2014). *Confirmatory factor analysis for applied research*. Guilford Publications.

Brunson, Emily K, Bettina Shell-Duncan, and Matthew Steele (2009). "Women's autonomy and its relationship to children's nutrition among the Rendille of northern Kenya". *American Journal of Human Biology: The Official Journal of the Human Biology Association* 21.1, pp. 55–64.

Carlson, Gwen J., Katarzyna Kordas, and Laura E. Murray-Kolb (2015). "Associations between women's autonomy and child nutritional status: A review of the literature". *Maternal & Child Nutrition* 11.4, pp. 452–482.

Castle, S E (1993). "Intra-household differentials in women's status: Household function and focus as determinants of children's illness management and care in rural Mali". *Health Transition Review* 3.2, pp. 137–57.

CSA and ICF (2000-2016). "Ethiopia Demographic and Health Survey 2000, 2011, 2015 and 2016. Addis Ababa, Ethiopia, and Rockville, Maryland, USA: Central Statistical Agency (CSA) and ICF International".

Currie, Janet and Douglas Almond (2011). "Human capital development before age five". *Handbook of Labor Economics*. Vol. 4. Elsevier, pp. 1315–1486.

DeCharms, R (1968). "Personal causation". *New York: Academic Press*.

Deci, Edward L and Richard M Ryan (2000). "The 'What' and 'Why' of goal pursuits: Human needs and the self-determination of behavior". *Psychological Inquiry* 11.4, pp. 227–268.

Demelash, Habtamu, Dabere Nigatu, and Ketema Gashaw (2015). "A case-control study on intimate partner violence during pregnancy and low birth weight, Southeast Ethiopia". *Obstetrics and Gynecology International* 2015.

Deyessa, Negussie, Y Berhane, Maria Emmelin, Mary C Ellsberg, Gunnar Kullgren, and Ulf Högberg (2010). "Joint effect of maternal depression and intimate partner violence on increased risk of child death in rural Ethiopia". *Archives of Disease in Childhood* 95.10, pp. 771–775.

Duflo, Esther (2003). "Grandmothers and granddaughters: Old-age pensions and intrahousehold allocation in South Africa". *The World Bank Economic Review* 17.1, pp. 1–25.

Dyson, Tim and Mick Moore (1983). "On kinship structure, female autonomy, and demographic behavior in India". *Population and development review*, pp. 35–60.

Egata, Gudina, Yemane Berhane, and Alemayehu Worku (2014). "Predictors of acute undernutrition among children aged 6 to 36 months in east rural

Ethiopia: A community based nested case-control study". *BMC Pediatrics* 14.1, p. 91.

Fafchamps, Marcel, Bereket Kebede, and Agnes R Quisumbing (2009). "Intrahousehold welfare in rural Ethiopia". *Oxford Bulletin of Economics and Statistics* 71.4, pp. 567–599.

Garoma, Sileshi, Mesganaw Fantahun, and Alemayehu Worku (2012). "Maternal intimate partner violence victimization and under-five children mortality in Western Ethiopia: A case–control study". *Journal of Tropical Pediatrics* 58.6, pp. 467–474.

Gelman, Andrew and Eric Loken (2013). *The garden of forking paths: Why multiple comparisons can be a problem, even when there is no 'fishing expedition' or 'p-hacking' and the research hypothesis was posited ahead of time*. Tech. rep. Mimeo, Department of Statistics, Columbia University.

Goetz, Anne Marie et al. (1997). *Getting institutions right for women in development*. Zed Books.

Gupta, Kamla and P Princy Yesudian (2006). "Evidence of women's empowerment in India: A study of socio-spatial disparities". *GeoJournal* 65.4, pp. 365–380.

Headey, Derek (2014). *An analysis of trends and determinants of child undernutrition in Ethiopia, 2000-2011*. ESSP working papers 70. IFPRI.

Hossain, SM Moazzem, Arabella Duffield, and Anna Taylor (2005). "An evaluation of the impact of a US $60 million nutrition programme in Bangladesh". *Health Policy and planning* 20.1, pp. 35–40.

Imai, Katsushi S, Samuel Kobina Annim, Veena S Kulkarni, and Raghav Gaiha (2014). "Women's empowerment and prevalence of stunted and underweight children in rural India". *World Development* 62, pp. 88–105.

Iversen, Vegard and Richard Palmer-Jones (2018). "All you need is cable TV?" *The Journal of Development Studies*, pp. 1–21.

Jejeebhoy, Shireen J. and Zeba A. Sathar (2001). "Women's autonomy in India and Pakistan: the influence of religion and region". *Population and Development Review* 27.4, pp. 687–712.

Kabeer, Naila (1999). "Resources, agency, achievements: Reflections on the measurement of women's empowerment". *Development and Change* 30.3, pp. 435–464.

Kishor, Sunita (2000). "Empowerment of women in Egypt and links to the survival and health of their infants". *Harriet B. and Sen, Gita (eds.). Women's Empowerment and Demographic Processes: Moving Beyond Cairo. New York/Oxford: Oxford University Press.*

Lépine, Aurélia and Eric Strobl (2013). "The effect of women's bargaining power on child nutrition in rural Senegal". *World Development* 45, pp. 17–30.

Leroy, Jef L (2011). *zscore06: Stata command for the calculation of anthropometric z-scores using the 2006 WHO child growth standards*. `http://www.ifpri.org/staffprofile/jef-leroy`. Accessed: 2019-01-25.

Lindstrom, David P and Betemariam Berhanu (2000). "The effects of breastfeeding and birth spacing on infant and early childhood mortality in Ethiopia". *Social biology* 47.1-2, pp. 1–17.

Mabsout, Ramzi (2011). "Capability and health functioning in Ethiopian households". *Social Indicators Research* 101.3, pp. 359–389.

Maitra, Pushkar (2004). "Parental bargaining, health inputs and child mortality in India". *Journal of Health Economics* 23.2, pp. 259–291.

Malhotra, Anju and Mark Mather (1997). "Do schooling and work empower women in developing countries? Gender and domestic decisions in Sri Lanka". 12.4, pp. 599–630.

Malhotra, Anju, Sidney Ruth Schuler, and Carol Boender (2005). "Women's empowerment as a variable in international development". *Measuring empowerment: Cross-disciplinary perspectives* 1.1, pp. 71–88.

Malhotra, Chetna, Rahul Malhotra, Truls Østbye, and SV Subramanian (2014). "Maternal autonomy and child health care utilization in India: Results from the National Family Health Survey". *Asia Pacific Journal of Public Health* 26.4, pp. 401–413.

Mason, Karen Oppenheim and Herbert L Smith (2000). "Husbands' versus wive' fertility goals and use of contraception: The influence of gender context in five Asian countries". *Demography* 37.3, pp. 299–311.

Muthén, Bengt and Tihomir Asparouhov (2012). "Bayesian structural equation modeling: A more flexible representation of substantive theory". *Psychological Methods* 17.3, pp. 313–335.

Muthén, L K and B O Muthén (1998-2012). "Mplus User's Guide. Seventh edition. Los Angeles, CA: Muthén and Muthén".

Pratley, Pierre (2016). "Associations between quantitative measures of women's empowerment and access to care and health status for mothers and their children: A systematic review of evidence from the developing world". *Social Science and Medicine* 169, pp. 119–131.

Pratley, Pierre and John Floyd Sandberg (2018). "Refining the conceptualization and measurement of women's empowerment in Sub-Saharan Africa using data from the 2013 Nigerian Demographic and Health Survey". *Social Indicators Research* 140.2, pp. 777–793.

Quisumbing, Agnes R and John A Maluccio (2003). "Resources at marriage and intrahousehold allocation: Evidence from Bangladesh, Ethiopia, Indonesia, and South Africa". *Oxford Bulletin of Economics and Statistics* 65.3, pp. 283–327.

Richardson, Robin A. (2018). "Measuring women's empowerment: A critical review of current practices and recommendations for researchers". *Social Indicators Research* 137.2, pp. 539–557.

Roushdy, Rania (2004). *Intrahousehold resource allocation in Egypt: Does women's empowerment lead to greater investments in children?* Working Papers 0410. Economic Research Forum.

Sandberg, John and Patrick Rafail (2013). "Measurement models of women's autonomy using the 1998/1999 India DHS". *Journal of Population Research* 30.4, pp. 367–381.

Schmeer, Kammi K (2005). "Married women's resource position and household food expenditures in Cebu, Philippines". *Journal of Marriage and Family* 67.2, pp. 399–409.

Sethuraman, Kavita, Richard Lansdown, and Keith Sullivan (2006). "Women's empowerment and domestic violence: The role of sociocultural determinants in maternal and child undernutrition in tribal and rural communities in South India". *Food and Nutrition Bulletin* 27.2, pp. 128–143.

Seymour, Greg and Amber Peterman (2018). "Context and measurement: An analysis of the relationship between intrahousehold decision making and autonomy". *World Development* 111, pp. 97–112.

Shroff, Monal R., Paula L. Griffiths, Chirayath Suchindran, Balakrishna Nagalla, Shahnaz Vazir, and Margaret E. Bentley (2011). "Does maternal autonomy influence feeding practices and infant growth in rural India?" *Social Science & Medicine* 73.3, pp. 447–455.

Smith, Lisa C, Usha Ramakrishnan, Aida Ndiaye, Lawrence Haddad, and Reynaldo Martorell (2003). *The importance of women's status for child nutrition in developing countries*. IFPRI Report 131. IFPRI Washington, D.C.

Staveren, Irene van and Olasunbo Ode bode (2007). "Gender norms as asymmetric institutions: A case study of Yoruba women in Nigeria". *Journal of Economic Issues* 41.4, pp. 903–925.

Svedberg, Peter (1990). "Undernutrition in Sub-Saharan Africa: Is there a gender bias?" *The Journal of Development Studies* 26.3, pp. 469–486.

Thomas, Duncan (1990). "Intra-household resource allocation: An inferential approach". *Journal of Human Resources*, pp. 635–664.

Thorpe, Sara, Kristin VanderEnde, Courtney Peters, Lauren Bardin, and Kathryn M Yount (2016). "The influence of women's empowerment on child immunisation coverage in low, lower-middle, and upper-middle income countries: A systematic review of the literature". *Maternal and Child Health Journal* 20.1, pp. 172–186.

Tibshirani, Robert (1996). "Regression shrinkage and selection via the lasso". *Journal of the Royal Statistical Society: Series B (Methodological)* 58.1, pp. 267–288.

Tosheno, Deneke, Yohannes Mehretie Adinew, Thilagavathi Thangavel, and Shimelash Bitew Workie (2017). "Risk factors of underweight in children aged 6–59 months in Ethiopia". *Journal of Nutrition and Metabolism* 2017.

Vaz, Ana, Pierre Pratley, and Sabina Alkire (2016). "Measuring women's autonomy in Chad using the relative autonomy index". *Feminist Economics* 22.1, pp. 264–294.

Vivalt, Eva (2018). "Specification searching and significance inflation across time, methods and disciplines". *Oxford Bulletin of Economics and Statistics*.

WHO et al. (2006). *WHO child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: methods and development*. Geneva: World Health Organization.

— (2010). *Nutrition Landscape Information System (NLIS) country profile indicators: interpretation guide*. Geneva: World Health Organization.

Appendix

TABLE C1: Child and parental characteristics

Variable	Mean	Std. Errors
Child characteristics		
Child age in month	24.760	0.362
Younger than 24 months (dv)	0.529	0.010
Child male (dv)	0.504	0.011
First born (dv)	0.216	0.009
Birth interval less than 24 months (dv)	0.094	0.006
Birth interval between 24 and 48 months (dv)	0.423	0.010
Birth interval more than 48 months (dv)	0.267	0.010
Breastfed within an hour after birth (dv)	0.655	0.012
Breastfed within a day after birth (dv)	0.224	0.010
Breastfed more than a day after birth (dv)	0.120	0.009
Never breastfed (dv)	0.003	0.001
Not wanted birth (dv)	0.729	0.010
Maternal characteristics		
Can't read (dv)	0.732	0.011
Can only read words (dv)	0.110	0.007
Can only read sentences (dv)	0.158	0.009
HH head: self (dv)	0.021	0.003
HH head: husband (dv)	0.943	0.005
HH head: other (dv)	0.036	0.004
Age when having first born baby	18.657	0.078
Age when first cohabitation	16.414	0.081
Child marriage (dv)	0.665	0.011
Age in years	28.844	0.160
Being a youngmom (dv)	0.576	0.011
Education: no (dv)	0.639	0.013
Education: primary (dv)	0.286	0.011
Education: secondary (dv)	0.055	0.006
Education: higher (dv)	0.020	0.003
Height	156.880	0.140
Work: currently working (dv)	0.286	0.012
Work: worked in the last 12 months (dv)	0.440	0.013
Being pregnant	0.101	0.007
Paternal characteristics		
Age in years	36.225	0.223
Education: no (dv)	0.475	0.013
Education: primary (dv)	0.393	0.012
Education: secondary (dv)	0.094	0.007
Education: higher (dv)	0.038	0.004
Different education years: mom-dad	-1.269	0.065
Age difference between dad and mom divided by average age	0.220	0.003
N	5119	

TABLE C2: Household characteristics

Variable	Mean	Std. Errors
Rural (dv)	0.861	0.012
Religion: orthodox (dv)	0.447	0.016
Religion: protestant (dv)	0.208	0.012
Religion: muslim (dv)	0.319	0.017
Religion: other (dv)	0.026	0.005
Wealth: poorest (dv)	0.201	0.012
Wealth: poorer (dv)	0.196	0.010
Wealth: middle (dv)	0.213	0.010
Wealth: richer (dv)	0.188	0.010
Wealth: richest (dv)	0.201	0.013
Average school years of parents	2.527	0.092
Head: male (dv)	0.965	0.004
Head: age	37.153	0.234
Number of children	3.463	0.053
Share of kids under five in family	0.269	0.003
Size	5.689	0.050
Drinking water: piped into dwelling (dv)	0.005	0.001
Drinking water: piped to yard/plot (dv)	0.065	0.006
Drinking water: piped to neighbor (dv)	0.010	0.003
Drinking water: public tap/standpipe (dv)	0.187	0.012
Drinking water: tube well or borehole (dv)	0.068	0.007
Drinking water: protected well (dv)	0.047	0.005
Drinking water: unprotected well (dv)	0.041	0.004
Drinking water: protected spring (dv)	0.185	0.012
Drinking water: unprotected spring (dv)	0.199	0.011
Drinking water: river/dam/lake/ponds/stream/canal etc. (dv)	0.185	0.012
Drinking water: rainwater (dv)	0.004	0.002
Drinking water: tanker truck (dv)	0.001	0.001
Drinking water: cart with small tank (dv)	0.002	0.001
Drinking water: bottled water (dv)	0.001	0.000
Drinking water: other (dv)	0.001	0.001
Drinking unclean water: unprotected well/spring/river etc. (dv)	0.425	0.015
Type of toilet: flush to piped sewer system (dv)	0.019	0.003
Type of toilet: ventilated improved pit latrine (dv)	0.008	0.002
Type of toilet: pit latrine with slab (dv)	0.064	0.005
Type of toilet: pit latrine without slab/open pit (dv)	0.435	0.014
Type of toilet: no facility/bush/field (dv)	0.451	0.015
Type of toilet: composting toilet (dv)	0.020	0.003
Type of toilet: bucket toilet (dv)	0.000	0.000
Type of toilet: hanging toilet/latrine (dv)	0.001	0.001
Type of toilet: other (dv)	0.001	0.001
Having no toilet facility (dv)	0.451	0.015
Floor is made of sand or dung (dv)	0.885	0.008
Roof is none, or made of thatch or mud (dv)	0.395	0.014
Cooking fuel: wood, dung,grass, or crop (dv)	0.918	0.007
Cooking fuel: charcoal (dv)	0.049	0.005
Cooking fuel: electricity, gas or fluid (dv)	0.032	0.003
Total pieces of large assets	0.678	0.027
No large asset (dv)	0.571	0.013
N	5119	

TABLE C3: PDS-LASSO regression models: variables selected for regressions with dependent variables of being underweight and WAZ

Panel A: Dependent variable	Being underweight					WAZ				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Independent variables:										
Autonomy	Y					Y				
Household decision-making		Y			Y		Y			Y
Non-acceptance of violence			Y		Y			Y		Y
Control over sex				Y	Y				Y	Y
Psu-wave FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
70 Unpenalised variables	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
LASSO selected variables:										
Average school years of parents	Y	Y	Y		Y	Y	Y	Y		Y
Region: Amhara # birthyear: 1993	Y	Y	Y	Y	Y					
Mom can't read at all # Wave 4	Y	Y			Y	Y	Y			Y
Mom can read sentence # breastfeed within a day	Y					Y				
Mom can't read at atll # wanted child	Y					Y				
Mom's education: higher # child's age	Y			Y	Y	Y			Y	Y
Dad's education: none # mom can read sentence	Y	Y			Y	Y	Y			Y
Mom can't read at all		Y			Y		Y			Y
Mom can read sentence # breastfeed more than a day		Y			Y		Y			Y
Mom's education: primary # breastfeed within an hour			Y		Y			Y		Y
Average school years of parents # child's age				Y	Y				Y	Y
Dad's education: primary # mom can read sentences				Y	Y				Y	Y
Dad's education: secondary # mom can't read at all				Y	Y				Y	Y
Mom's height						Y	Y	Y	Y	Y
First born kid # birthyear: 1997						Y	Y	Y	Y	Y
N					5119					

Note: This table reports the variables selected for models depicted in Panel A in Table 3.5. All models apply a *psu*-wave fixed-effects to the PDS-LASSO model. The unpenalised variables are birth year, birth month, rural, child's age in months, gender, first born, and birth interval.

TABLE C4: PDS-LASSO regression models: variables selected for regressions with dependent variables of being stunted and HAZ

Panel B: Dependent variable	Being stunted					HAZ				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Independent variables:										
Autonomy	Y					Y				
Household decision-making		Y			Y		Y			Y
Non-acceptance of violence			Y		Y			Y		Y
Control over sex				Y	Y				Y	Y
Psu-wave FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
70 Unpenalised variables	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
LASSO selected variables:										
Average school years of parents	Y	Y	Y		Y	Y	Y	Y		Y
Maternal height	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Mom can't read at all # wave 4	Y	Y			Y	Y	Y			Y
Mom can read sentence # breastfeed within a day	Y					Y				
Mom can't read at all # wanted child	Y					Y				
Mom's education: higher # child's age	Y			Y	Y	Y			Y	Y
Maternal height # wave 3	Y	Y	Y	Y	Y					
Dad education: no education # mom can read sentence	Y	Y			Y	Y	Y			Y
Mom can't read at all		Y			Y		Y			Y
Mom can read sentence # breastfeed more than a day		Y			Y		Y			Y
Mom's education: no education # breastfeed within an hour			Y		Y			Y		Y
Average school years of parents # child's age				Y	Y				Y	Y
Dad's education: primary # mom can read sentence				Y	Y				Y	Y
Dad education: secondary # mom can't read at all				Y	Y				Y	Y
N						5119				

Note: This table reports the variables selected for models depicted in Panel B in Table 3.5. All models apply a *psu*-wave fixed-effects to the PDS-LASSO model. The unpenalised variables are birth year, birth month, rural, child's age in months, gender, first born, and birth interval.

TABLE C5: Mechanism: heterogeneous effect

Dependent variable	Types of staple food child ate		
	(1)	(2)	(3)
Household decision-making	0.334***	0.281***	0.178**
	(0.069)	(0.081)	(0.084)
psu FE	Y	Y	Y
Child-level controls	Y	Y	Y
Mother-level controls		Y	Y
Household-level controls			Y
N	320		

Note: Robust standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The sample consists of children aged 24-59 months. All models apply a *psu*-wave fixed-effects. The dependent variable is the summary of types of staple food child ate, including fortified baby food, teff, lentils, and solid and semi-solid food, with a mean equals to 0.94. Child-level controls are the birth month, birth year, child's age, square of child's age, gender, birth interval, and wanted birth. Mother-level controls are literacy, relationship to household head, mother's age, square of mother's age, education level, work in the past week, and work in the past year. Household-level controls are religion, rural, mean of educational years of parents, household head's gender, household head's age, household head's age square, no of children, share of children under five in the family, household size, square of household size's, drinking unclean water, no toilet facility, no large asset, the floor is made of inferior materials, the roof is made of inferior materials, and type of cooking fuel.

TABLE C6: PDS-LASSO regression models: without unpenalised variables

Dependent variable	Being underweight (dv) (Mean=0.254)				
	(1)	(2)	(3)	(4)	(5)
Autonomy	-0.017*				
	(0.010)				
Household decision-making		-0.010			-0.004
		(0.009)			(0.009)
Non-acceptance of violence			-0.017**		-0.016*
			(0.008)		(0.008)
Control over sex				-0.011	-0.007
				(0.008)	(0.008)
N			5119		

Note: Region-wave-rural clustered standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All models apply a *psu*-wave fixed-effects to the PDS-LASSO model. In this table, we present results from models in which we do not unpenalise any variables. These coefficients of regressions and the mean of outcome variables are estimated using the weights from EDHS.

TABLE C7: PDS-LASSO regression models: heterogeneous results using restricted sample

Panel A: young child	Being underweight (dv) (Mean=0.253)			
(=1 if younger than 24 months)	(1)	(2)	(3)	(4)
Autonomy	-0.035**			
	(0.017)			
Autonomy × young child	0.029			
	(0.021)			
Household decision-making		-0.024*		
		(0.012)		
Household decision-making × young child		0.025		
		(0.016)		
Non-acceptance of violence			-0.040**	
			(0.018)	
Non-acceptance of violence × young child			0.041*	
			(0.023)	
Control over sex				-0.017
				(0.016)
Control over sex × young child				0.005
				(0.021)
N	4566			

Note: Region-wave-rural clustered standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All models apply a *psu*-wave fixed-effects to the PDS-LASSO model. A restricted sample is used here, excluding pastoral areas – Afar and Somali. The unpenalised variables are birth year, birth month, rural, child's age in month, gender, first born, and birth interval. These coefficients of regressions and the mean of outcome variables are estimated using the weights from EDHS.

TABLE C8: PDS-LASSO regression models: when women have highest autonomy if they make decision jointly with husband

Dependent variable	Being underweight (dv) (Mean=0.254)			
	(1)	(2)	(3)	(4)
Autonomy	-0.016*			
	(0.009)			
Household decision-making		-0.009		
		(0.009)		
Non-acceptance of domestic violence			-0.016	
			(0.011)	
Control over sex				-0.012
				(0.008)
N			5119	

Note: Region-wave-rural clustered standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All models apply a *psu*-wave fixed-effects to the PDS-LASSO model. New Bayesian CFA model is applied when the decision-making is scored in an alternative way, i.e., women making jointly decision has a higher score than women making decision alone. Therefore the regressions apply the new measure of autonomy. The unpenalised variables are birth year, birth month, rural, child's age in month, gender, first born, and birth interval. These coefficients of regressions and the mean of outcome variables are estimated using the weights from EDHS.

TABLE C9: PDS-LASSO regression models: restricted sample to women having only one child under five

Dependent variable	Being underweight (dv) (Mean=0.255)			
	(1)	(2)	(3)	(4)
Autonomy	-0.024*			
	(0.014)			
Household decision-making		-0.022*		
		(0.011)		
Non-acceptance of domestic violence			-0.008	
			(0.014)	
Control over sex				-0.022**
				(0.010)
N	3282			

Note: Region-wave-rural clustered standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All models apply a *psu*-wave fixed-effects to the PDS-LASSO model. The sample is restricted to women having only one child under five. The unpenalised variables are birth year, birth month, rural, child's age in month, gender, first born, and birth interval. These coefficients of regressions and the mean of outcome variables are estimated using the weights from EDHS.

FIGURE D1: Conventional CFA model with standardised factor-loadings

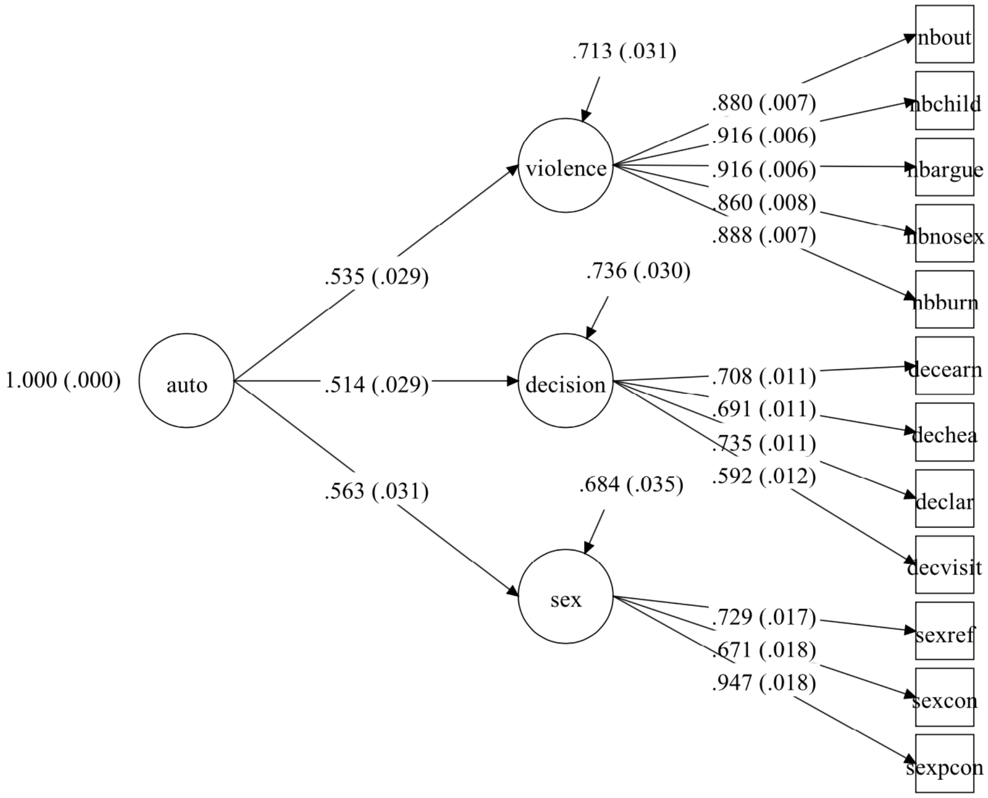
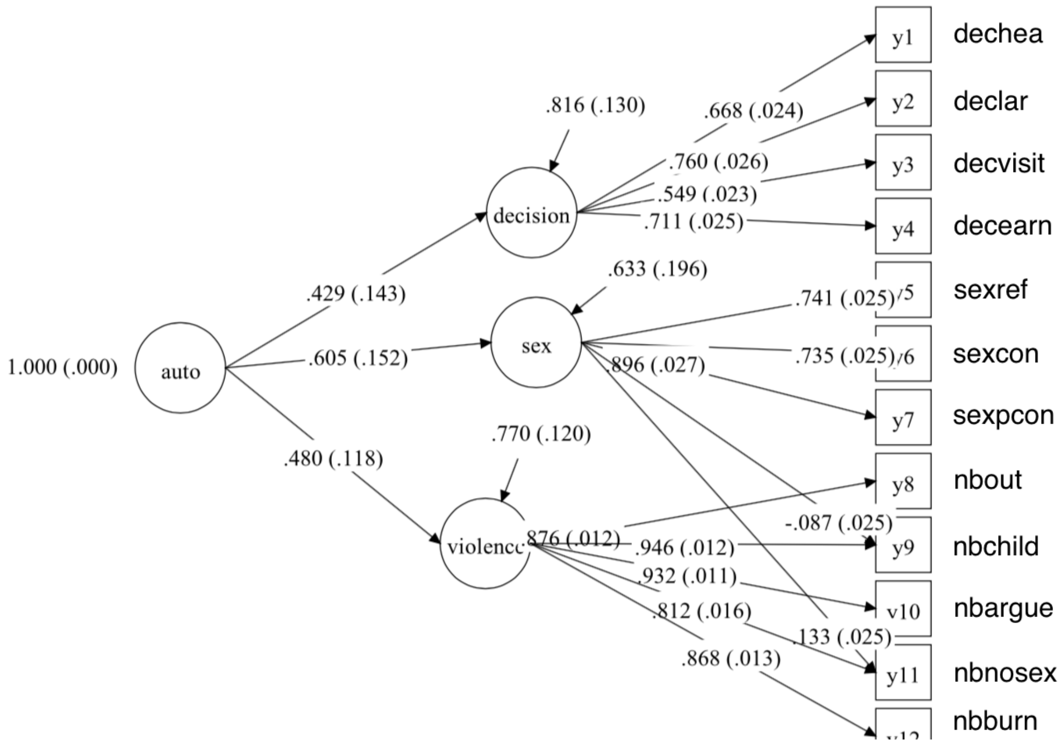


FIGURE D2: Bayesian CFA model with standardised factor-loadings, allowing for cross-loadings



Chapter 4

Do early life shocks constrain returns to human capital investment?
Evidence from a large-scale sanitation programme in India

Abstract

This paper explores whether the productivity of investments in early childhood is constrained by in-utero adverse shocks, which could potentially decrease returns on subsequent investments in child development. To answer this question, we exploit two exogenous sources of variation conditional on variation in district and time - fetal nutrition input shocks proxied by local rainfall deviations and exposure to a sanitation campaign at birth - to investigate the interaction between endowments and investment. We use a difference-in-differences design to estimate impacts on a measure of cognitive skills of children aged 8-11 in rural India. We find that positive rainfall in utero and India's Total Sanitation Campaign (TSC) in the first year of life increase test outcomes. However, we do not find a significant heterogeneous effect of TSC by previous exposure to positive rainfall in utero, meaning that in-utero adverse shocks do not constrain or boost the returns on investments in the next stage of child development.

Keywords: Early-life Influences, Dynamic Skill Formation, Sanitation Programme.

JEL classification: J24, O15, I38.

4.1 Introduction

Fetal and early-life conditions, such as nutritional inputs and environment, play an important and long-lasting role in children's lives (Almond and Currie, 2011; Currie and Almond, 2011; Heckman, 2007). Parents and policy-makers have been trying to eliminate the deficits in child development caused by adverse early-life events by providing disadvantaged children with more resources and designing interventions in the later life of children (Almond and Mazumder, 2013; Aizer et al., 2016). However, the question of whether the detrimental effects of poor circumstances and insufficient investments in early-life could be alleviated or undone by these interventions is not sufficiently answered.

A growing literature has begun to learn the technology of human capital production, especially assessing the question of whether the return to investment responds to baseline capabilities homogeneously or heterogeneously. If heterogeneously, one would like to know whether the baseline stock of human capital is restricting the return of investment or it is boosting it. The hypothesis of *dynamic complementarities* proposed by Heckman (2007) suggests the former case, where investments in human capital would be more productive when previous investments or baseline capabilities are higher. According to the theory of dynamic human capital formation, timing plays a significant role in correcting for deficiency in human capital development (Cunha et al., 2010; Heckman and Mosso, 2014). As the child ages, the degree of static substitutability between investments and baseline capabilities reduces. To earn the largest return to investments in children, investing very early in childhood would be a better choice than in later childhood (Cunha et al., 2006; Doyle et al., 2009; Heckman, 2006). What is even worse is that when complementarities kick in, investments might be completely inefficient when the stocks of human capital are very low (Cunha et al., 2006; Doyle et al., 2009; Heckman, 2006; Conti and Heckman, 2014; Heckman and Mosso, 2014). However, at which point in time the baseline stocks of skills start to constrain the returns to investment (i.e. the occurrence of *dynamic complementarities*) is still an open question.

The empirical challenge of pinning down the causal impact of *dynamic complementarities*, a crucial element to understand human capital formation, is addressed in Almond and Mazumder (2013) as follows, "(a) exogenous variation in the baseline stock and (b) exogenous variation in subsequent investment... this may be asking for lightning to strike twice: two identification strategies affecting the same cohort but at adjacent developmental stages". In response to this estimation challenge, following studies including Adhvaryu et al. (2018), Duque et al. (2018), Gunnsteinsson et al. (2018), Johnson and Jackson (2017), Malamud et al. (2016), and Rossin-Slater and Wüst (2016), this paper combines exogenous variation in rainfall shocks with variation in sanitation conditions using difference-in-differences estimation to show whether the return to an improved sanitation environment would differ with baseline stock of human capital.

Specifically, the first variation is generated by the exogenous rainfall shock. As the occurrence of good or bad rainfall is a quasi-random event across districts, together with the randomness in the birth time, it provides a natural experiment.[1] In a typical rainfed country like India, rainfall failure decreases agriculture production and agricultural wages heavily (Jayachandran, 2006; Colmer, 2018), and human capital investment in children is highly dependent on wages (Jacoby and Skoufias, 1997; Jensen, 2000; Thomas et al., 2004).

[1]Recently economists have exploited natural experiments which are quasi-random in nature to identify the causal impact of early life circumstances on later-life outcomes, such as famine (Chen and Zhou, 2007; Dercon and Porter, 2014), pandemics (Almond, 2006; Banerjee et al., 2010), and extreme rainfall shocks (Maccini and Yang, 2009; Shah and Steinberg, 2017).

Since rural households in developing countries are highly reliant on agriculture income to support their living, the literature argues that droughts would adversely impact on a child's early-life nutrition input and therefore induce early disadvantage in endowment and damage human capital development (Maccini and Yang, 2009; Rocha and Soares, 2015; Rosales, 2014; Krutikova and Lilleør, 2015). For example, in the context of India, exposure to deficient rainfall in early-life discourages children's cognitive development (Shah and Steinberg, 2017) and stunts children's health outcomes (Kumar et al., 2016).

The second main source of variation is from a government sanitation programme called Total Sanitation Campaign (TSC). This programme was introduced in 2000, promoting the construction of low-cost pit latrines in rural areas, with *ex post* financial support to villages. In India, one of the highest environmental risks is the unsafe disposal of human faeces, where 53 per cent of households defecate in the open (UNICEF and WHO, 2014). After the programme was implemented, Spears (2012a), Kumar and Vollmer (2013), and Spears and Lamba (2016) find evidence that the programme successfully reduces childhood diarrhoea and infant mortality rate, and increases height and cognitive skills in rural India. Following Spears and Lamba (2016), we use the difference-in-differences(DID) estimation strategy to show that children born in an environment with higher sanitation intensity display better outcomes in later life.

We combine three main sources of data: firstly, we use the India Human Development Survey 2011 (IHDS-II) to gather information of children in rural areas, including test outcomes and anthropometric measurements, birth information, parents, households, and village characteristics; secondly, we merge the survey with rainfall data from University of Delaware to identify adverse rainfall shocks using birth information; lastly, we merge the survey data with TSC administrative data from the Indian government. The availability of test outcomes restricts the sample to children aged 8 to 11.

Our main interest is to leverage the combination of these two exogenous sources of variation in investments, in order to understand how baseline stocks of human capital and investments interact in producing capabilities in later childhood. We find that the estimates of interaction between two investments are positive yet non-significant, suggesting that children with higher endowments, compared to low-endowed children, do not benefit significantly larger from a better sanitation environment in generating higher test scores in later childhood. The overall result is consistent with one recent working paper, which suggests no evidence of dynamic complementarities of investments (Malamud et al., 2016).

Our study contributes to the literature in three ways. First of all, our empirical context is of high potentiality for external validity, as the shortage in rainfall is one of the most common shocks that low-income households experience in developing countries. We confirm the long-term effect of adverse rainfall on the human capital development in agricultural countries (Adhvaryu et al., 2018; Dercon and Porter, 2014; Maccini and Yang, 2009; Paxson, 1992; Shah and Steinberg, 2017). Second, this study complements the evidence showing TSC's impact on child development (Kumar and Vollmer, 2013; Spears, 2012a; Spears and Lamba, 2016) by discussing its heterogeneous effect across child endowments. Last but not least, we shed new light on the understanding of the complex process of human capital formation by studying interaction of a sanitation programme at birth with child endowments, while most of the existing studies focus on school interventions and use investments in later childhood; also we provide suggestive evidence that *dynamic complementarities* kick in at a very early age.

The rest of the paper comprises of a literature review, theoretical framework, introduction of the context of TSC, data and variable constructions. The econometric method is depicted, followed by results and conclusions.

4.2 Literature review

In the epidemiological literature, the fetal origins hypothesis (Barker, 1995) demonstrates that poor in utero nutrition could trigger chronic conditions of adult health. Economists follow this route and investigate a larger range of shocks in utero (such as food availability and disease environment) and find long-lasting detrimental effects on non-health outcomes, including cognitive development, life expectancy and adult earnings (see e.g. Almond, 2006; Almond et al., 2009; Banerjee et al., 2010).

Besides the well-established fetal origins literature, a growing and large body of studies have provided evidence that early childhood (postnatal) circumstances, including any shocks in weather, nutrition, disease, and pollution that children experience after birth, can have long-life consequences in human capital (Currie, 2009; Currie and Almond, 2011; Currie and Vogl, 2013; Heckman, 2006, 2007).

One strand of literature has found that adverse weather shocks in utero and early life impact negatively on future outcomes of children in agricultural countries, through the channel of nutritional intake (see e.g. Krutikova and Lilleør, 2015; Maccini and Yang, 2009; Rosales, 2014; Shah and Steinberg, 2017). In developing countries, the profitability of agricultural investments and agricultural wages is diminished hugely by rainfall failure (Colmer, 2018; Jayachandran, 2006; Rosenzweig and Binswanger, 1992); a decrease in wages, meanwhile, could lessen human capital investment (Jacoby and Skoufias, 1997; Jensen, 2000; Thomas et al., 2004) and nutritional input through decreased food consumption (Kumar et al., 2016; Rocha and Soares, 2015). In return, this could prohibit human capital development of children and generate undesirable educational and economic outcomes (Maccini and Yang, 2009; Shah and Steinberg, 2017).

Current empirical evidence from developing countries has suggested that rainfall shocks in early-life have an impact on cognitive and non-cognitive development in later life. These studies have also exploited the timing of exposure and suggested that the effect is significant mainly during in utero and in birth year. Maccini and Yang (2009) use Indonesia Family Life Survey (IFLS) to examine what role early-life rainfall shocks play on later-life outcomes, such as health and education, in adulthood in Indonesia with a district and birth year fixed-effects model. They find a positive effect of rainfall at birth on the completed grade of schooling, yet this effect is only found for girls. Using a village of residence and birth year fixed-effects model, Rosales (2014) exploits the occurrence of El Niño in Ecuador, which induces an exogenous variation in contemporaneous household income and consumption and hence endowments, and finds that in-utero exposure to rainfall shocks has a persistent impact on children's cognitive outcomes in later childhood measured by the Peabody Picture Vocabulary Test scores. However, there is no effect found from exposure in the first year of life. Krutikova and Lilleør (2015) study the impact of rainfall on another dimension of human capital (i.e. non-cognitive skills) measured by core self-evaluation in adulthood using a sibling-fixed effects model in Tanzania. They investigate the timing of exposure to rainfall deviation in early life and find that the effect is only significant during in utero and not in the first two years of life. In the context of India, Shah and Steinberg (2017) use more than 2 million rural primary school children in India surveyed by an educational achievement called the Annual Status of Education Report (ASER). The authors find that rainfall shocks in utero and the first three years of life have a significant impact on test scores and school enrolment using a household fixed-effects model. Also, their results suggest that the in utero effects are larger for girls.

Another devastating environment causing gaps in childhood development in low-income countries is poor sanitation, as Grantham-McGregor et al. (1999) regard it as the top barrier for children to accumulate human capital. Both the epidemiological literature and economic studies have shown that open defecation, which generates a disease environment, has a large effect on early-life health, nutrition, and cognitive development. There are two mechanisms recognised: diarrhoea (Checkley et al., 2008; Guerrant et al., 1999) and tropical or environmental enteropathy (Humphrey, 2009; Mondal et al., 2011; Petri et al., 2008).

Some empirical studies provide indicative evidence in the correlation between contemporaneous sanitation environment and human capital, using cross-sectional data. For example, Spears (2012b) finds that contemporaneous sanitation practice of households might be an important factor in explaining better outcomes of children in height and cognitive achievement, using an OLS model and data from India Human Development Survey 2005 (IHDS-I). Vyas et al. (2016) show that better sanitation explains the increase in height in Cambodia between 2005-2010 and that externalities of open defecation are the main cause for worse outcomes rather than the child behaviour of defecation itself. Most enlightening is that it is the open defecation at community-level that matters for a child's human capital rather than the household-level. This sheds light on the important role for public policy in improving sanitation environment on a large-scale.

There has been a global focus on improving the disease environment by having affordable interventions for adequate sanitation, in order to fulfil children's developmental potential in the developing world (Cumming and Cairncross, 2016; Mbuya and Humphrey, 2016). For instance, in India, where over half of the population defecate openly, one national programme called the Total Sanitation Campaign (TSC), with an aim to construct low-cost pit latrines in rural areas, was started in 2001 and is the focus of this study.

Although a few studies using small-scale randomised controlled trials have shown that TSC has had no significant impact on child health (Clasen et al., 2014; Patil et al., 2014), some other studies using a larger scale of data find effects on human capital of the sanitation programme. Using the third round of District Level Household Survey 2007 (DLHS-III) in India and implementing propensity score matching, Kumar and Vollmer (2013) find that the probability of children under the age of five contracting diarrhoea drops by 2.2 percentage points when they have access to improved sanitation. The authors also investigate the heterogeneous effect and find that the relationship holds particularly for boys and children from better socioeconomic status families. Exploiting a cluster-randomised trial in rural Mali, Pickering et al. (2015) study the effect of a community-led sanitation intervention and find that the increased access to toilet improves child health, especially for those under the age of two. Using a cross-sectional data for city Gwalior, Madhya Pradesh, India, Augsburg and Rodríguez-Lesmes (2018) study the effect of a sanitation intervention called FINancial Inclusion Improves Sanitation and Health (FINISH), which is run by a voluntary organisation. They examine the causal effects of the sanitation environment at village-level by instrumenting it by the collected local raw material construction prices. They find that sanitation coverage in the first years of life plays a significant role in height growth for children under five and that girls respond more positively to the improved disease environment.

Using the difference-in-differences strategy, Spears (2012a) was the first to identify the causal effect of rural Total Sanitation Campaign (TSC) on mortality rate and height in India. Using DLHS-III, he finds that the mortality rate has declined; using IHDS-I, he also shows that children under five grow taller if they are exposed to higher accumulated TSC latrines in the first year of their life. Using the same econometric strategy, Spears and Lamba (2016) fill the gap in the literature in linking sanitation and cognitive achievement. Using ASER data, the study shows that the TSC intensity exposed by six-year-old children in their first year of life increases their test outcomes. Moreover, the results suggest that the sanitation-cognition gradient is steepest in the first year of life for exposure, as there is no effect found when exposure to the TSC at other ages excepts age zero.

Following Spears and Lamba (2016), we will exploit whether the effect of sanitation environment is stronger for children with higher endowments, in which we consider rainfall shocks as an in-utero natural experiment that induced variation. Although a few recent studies have shown the heterogeneous return on school investments with respect to child endowments (Adhvaryu et al., 2018; Aizer and Cunha, 2012; Duque et al., 2018; Johnson and Jackson, 2017; Malamud et al., 2016), we are the first to examine this in the context of sanitation investments.

There is a small but growing literature studying the interactive effect between investments and endowments, despite the identification obstacles raised by difficulties in finding an exogenous investment to interact with another source of exogenous variation in endowments (Almond and Mazumder, 2013). However, the results are conflicting, as some finds evidence on complementarities (Aizer and Cunha, 2012; Johnson and Jackson, 2017); some on substitution (Rossin-Slater and Wüst, 2016) or even mitigation (Adhvaryu et al., 2018; Gunnsteinsson et al., 2018); some find little evidence on interaction (Malamud et al., 2016); and lastly, Duque et al. (2018) find that the timing of the subsequent investments matters for the interaction effects, as complementarites can only be found if investments happen in early childhood.

First of all, Johnson and Jackson (2017) show evidence of *dynamic complementarities* between investments. They use a difference-in-differences instrumental variables model to study the interaction between policy-induced changes in preschool (Head Start) spending and school-finance-reform-induced changes in public K12 school spending during childhood in the US. They find that the benefits of Head Start are larger when followed by the access to better-funded public K12 schools using outcomes such as high school graduation, years of education, and adult wages. Similarly, Aizer and Cunha (2012) exploit exogenous variation in preschool investments from the rollout of Head Start to study the interactive effect between these investments and stocks of early human capital measured by Bailey test score at the age of eight months. They show that preschool investments have a larger effect for children with higher test scores in infancy on a measure of cognitive skill at age four.

On the other hand, some find evidence of substitution. Using a difference-in-differences design, Rossin-Slater and Wüst (2016) test the interaction effect between two exogenous interventions in Denmark - nurse home visiting and childcare - on adult educational attainment, earnings, and survival beyond the age of 65. They reveal that a child who did not experience nurse home visiting gained larger benefits from childcare. They find a negative interaction effect, as nurse home visiting in the birth year reduces the return to the preschool programme at the age of three, suggesting that substitutability in health-related inputs in the first three years of life exists.

Several studies find evidence on mitigation in that return to investments may be higher for children exposed to adverse shocks. Gunnsteinsson et al. (2018) use a cluster-randomised controlled trial of vitamin A supplementation at birth and find that the added value of the programme increases for children exposed to tornado shock in utero and first three months of life in Bangladesh, using anthropometric measurements and incidence of severe fevers at 0-6 months as outcomes. They also find effects of tornado and protective effects of the vitamin A programmes are both larger for boys. However, restricted by the data, the evidence could only support a short-term effect of substitutability, failing to show the long-term effect of the interaction. Adhvaryu et al. (2018) study a large-scale randomised conditional cash transfer programme called PROGRESA in Mexico, focusing on its education component which gives cash payments conditional on children going to school during school years, and its interaction with exposure to rainfall shock around the time of birth. They use outcomes including educational attainment for children aged 12-18 and higher education enrollment and employment outcomes for children aged 18. They observe one additional year exposed to the programme could mitigate the discrepancy caused by the adverse rainfall.

Meanwhile, few find little evidence on interactive effects between shocks and investments. Using the combination of difference-in-differences and regression discontinuity design, Malamud et al. (2016) combine access to abortion and access to better schools and find a weak negative interaction effect on a school-leaving exam in Romania.

Lastly, Duque et al. (2018) are the first to exploit difference the timing of investment makes on the interaction effects. They use large-scale data from Colombia and a difference-in-differences framework combined with a regression discontinuity design to study the interaction effect between a conditional cash transfers (CCT) programme, which aims at stimulating investments in children's health and education, and early-life rainfall shocks. They find that when the CCT occurs in the early childhood, the return on test outcomes is higher for children exposed to positive rainfall shock, while the interaction is zero or smaller if CCT occurs in adolescence.

Among all of these studies, the majority have looked at the investments in developed countries with only a few in developing countries (Adhvaryu et al., 2018; Duque et al., 2018; Gunnsteinsson et al., 2018; Malamud et al., 2016), while understanding of skill formation is particularly important in the third world.

4.3 Theoretical framework

Cunha and Heckman (2007) specify that skill formation is a life-cycle process and that one characteristic of skill formation is *dynamic complementarity*, which means investments in early-life could raise the productivity of later investments. We modify their skill formation using cognitive skills as the outcome variable using the following technology:

$$\theta_{t+j,c} = f_{t,c}(h, \theta_t, I_t) \tag{4.1}$$

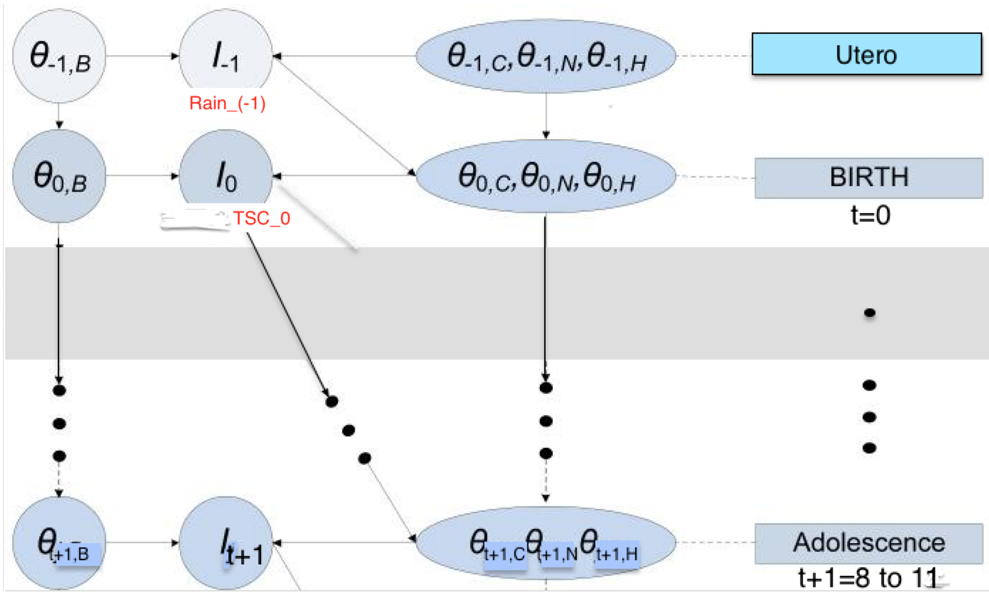
where θ_{t+j} is a vector of later skills measured at time $t + j$; c stands for cognitive skills; h represents parental characteristics; θ_t is a vector of former skills measured at time t; and I_t stands for investments in child skill during period t.

Equation 4.1 shows that skill itself at time t could be an input factor of the human capital production function. *Dynamic complementarity* means that the investment at time t will be more productive if the baseline skill θ_t is higher:

$$\frac{\partial^2 \theta_{t+j,c}}{\partial \theta_t \partial I_t} = \frac{\partial^2 f_{t,c}(h, \theta_t, I_t)}{\partial \theta_t \partial I_t} > 0 \tag{4.2}$$

In this life-cycle skill information process, $\theta_{t+j,c}$ is determined by both θ_t and I_t, while these two terms could interact with each other, generating a heterogeneous return of I_t. In this study (using Figure 4.1 as an illustration), the

FIGURE 4.1: Diagram of technology of capability formation



baseline skill is θ_0, which is driven by rainfall shock in utero I_{-1}. The investment in second stage I_0 is the TSC programme in the first year of life, which improves the sanitation environment. The cognitive skills $\theta_{t+j,c}$ are measured at the time when a child is aged from 8 to 11.

Therefore, in this study, our interest is to study the sign of the following term:

$$\frac{\partial^2 \theta_{t+j,c}}{\partial \theta_0 \partial I_0} = \frac{\partial^2 f_{0,c}(h, \theta_0, I_0)}{\partial \theta_0 \partial I_0} \tag{4.3}$$

If this term is larger than zero (positive interaction term between the stock of skill θ_0 and the investment I_0), it suggests *dynamic complementarities*; if this is smaller than zero (a negative interaction term), it implies *dynamic substitutabilities*.

As we do not have information on non-cognitive skills and health, we are assuming there is no cross-productivity between cognitive skills and other dimensions of human capital.

4.4 Data

4.4.1 The TSC programme

4.4.1.1 Description of programme

In 2001 (i.e. the beginning of the time period of this study) the Indian government launched a countrywide programme called the Total Sanitation Campaign (TSC) in rural areas, with the aim of promoting the construction of pit latrines in rural areas. The main goal of TSC was to end open defecation by 2017. This is a low cost and effective method to attain safe disposition of human excreta, taking around U.S. $30-$50 to build one pit latrines for ten rural people. In the first ten years of the programme, the open defecation in India has fallen from 63.3% to 53.1%. We will look at the effects of the first four years of TSC implementation in the time period of 2001-2004 on achievement in 2011/12, as the cohorts of 2001-2004 are aged from 8 to 11.[2]

TSC has successfully decreased the level of open defecation, though it did not eliminate open defecation completely (Barnard et al., 2013). By providing an *ex post* monetary incentive to local village government when the village is verified to be open defecation free in each financial year, the programme has successfully stimulated constructions of latrines and promoted latrine use.[3] Villages are divisions of districts; districts are divisions of states.

Although this programme of the central Indian government is designed to improve rural sanitation across the whole country, there are notable policy variations across the states and districts (Links, 2008). These variations stem

[2]Children who are born after 2004 will be younger than 7 in the survey IHDS-II, so that they are not administered in the tests.

[3]Spears (2012a) shows that this monetary incentive only played a role in stimulating TSC constructions until it was introduced.

from the distinct approaches and strategies as a result of that Indian state, mid-level and local government officials have autonomous power.[4] Moreover, at micro-administration level, village leaders are actually the ones who are in charge of the progress of constructing latrines in rural areas. That is to say, local village leaders will play a big role in the achievement of eliminating open defecation, which is highly dependent on their wills and social power. With the awareness of the potential correlation among policy outcomes and socio-economic variables suggested by a large body of political economy literature, in Section 4.6.4.2 we will check whether the within-district implementation of TSC is correlated with any within-district trends in a series of observable socio-economic characteristics.

4.4.2 TSC data

The TSC implementation records are collected by the government of India at the district level for each financial year. The year-wise district-level TSC achievements are public on the government website.[5] The data covers 632 districts in the time period 1999-2013, with records of zero in 1999-2000 since the programme was not implemented until 2001. It is obvious that since 2003, in the year when the *ex post* monetary incentives were announced by the central government, the number of construction of latrines has increased.[6]

[4]Links (2008) documents that different states and districts collaborated with different organizations such as the Rural Development Department, Public Health Engineering Department and NGOs.

[5]See https://data.gov.in/catalog/nirmal-bharat-abhiyan-year-wise-district-level-achievements. The data was downloaded on March 1, 2016.

[6]The central government awards monetary prizes to villages where open defecation free status is achieved in that year. For each financial year, data of open defecation at village-level is reported and ranked at the end of the year, and villages on the top of the list are awarded. Therefore, the leaders of villages would not have knowledge about whether they will win a prize at the time they make their decisions on constructing latrines. Therefore it is an *ex post* incentive.

4.4.2.1 TSC intensity variable

As to measure one of the key interest variables, we follow Spears and Lamba (2016)'s fashion in constructing the intensity of the programme. The raw TSC data reports latrine *counts* for each district in each year, so we will accumulate the numbers of TSC to get the total number of construction of latrines by each year. In India, the size of districts varies, and the density matters for disease environment. Therefore, to better measure the effect of the programme, we compute the district-level TSC intensity, i.e. the cumulative number of TSC latrines divided by the rural population in the 2001 census. It is shown as follows,

$$TSC_{dt} = \frac{\sum_{2001}^{t} \text{TSC latrines counts}_{dk}}{\text{rural population}_{d}} \times 100 \tag{4.4}$$

where t is the year being studied, and 2001 is the year when the programme started; d denotes for district. As shown in Equation 4.4, the nominator is the cumulative latrines built in the time period between 2001-t. The denominator is the district-level rural population collected from the Census of India 2001.[7] To ease the interpretation of the results, considering that the intensity is rather smaller in the first four years of the implementation of TSC, we multiply the intensity by 100. For example, one TSC intensity means that for every 100 persons there is a TSC latrine. In the robustness check, we will provide results using a new measurement of TSC intensity by applying a time-varying population for each year, using both the census of 2001 and 2011 to predict the annual growth rate of population.

Following Spears and Lamba (2016) to study the effect of TSC, we merge the constructed TSC intensity variable to children who are born in years 2001-2004 as a human capital investment or a positive environmental shock at birth,

[7] See http://censusindia.gov.in/. The data was downloaded on March 1, 2016.

as represented by I_0 in Equation 4.3, using information of districts of residence and birth year.

4.4.3 Rainfall data

The rainfall data is taken by the monthly rainfall data collected by the University of Delaware in order to construct rainfall shocks within districts.[8] The data consists of precipitation information covering all India in the period between 1900-2008, and we choose to use 34-year data from 1975-2008. The spatial resolution we use is 0.5×0.5 latitude by longitude so that the data can be accessed by grids. To match the data to districts, we choose the closest point on the grid to the geographical centre of the district, following Shah and Steinberg (2017).

4.4.3.1 Rainfall shock variable

As discussed earlier, a country like India is mostly rainfed, and almost every rural household depends on agriculture for their income. Therefore, rainfall fluctuations will cause potential crop failure and income shock for rural households (Guiteras, 2009; Jayachandran, 2006; Colmer, 2018). Specifically, it is the drought that engenders the most damaging effect on crop yield in predominantly rainfed areas in India, as approximately two-thirds of the cultivated land depends highly on monsoons (Auffhammer et al., 2012). Unlike some countries in South America where crops are vulnerable to extreme weather (droughts or floods), crops failure in South Asia is mostly caused by droughts.

We follow the definition of drought by India Meteorological Department's (IDM) (Attri and Tyagi, 2010) to identify the year of a drought year. The India

[8] The data can be found at http://climate.geog.udel.edu/ climate. The data was downloaded on March 1, 2016.

Meteorological Department defines a drought occurs when the monsoon rainfall, which is the rain between June and September, falls below 75 per cent of the long-term mean monsoon rainfall. The long-term mean is calculated using the 34-year rainfall from 1975 to 2008. Therefore, we construct a dummy variable to portray a year with or without rainfall shocks. In order to ease the interpretation, we use 'normal rainfall' to represent a year without the occurrence of drought. The dummy variable equals one if there is no drought during the year; it is equal to zero, otherwise.

To check the sensitivity to the cut-off of the definition of 'normal rainfall', we try varying the cut-off between 70% to 80%. Table 4.1 shows the means of 'normal rainfall' in utero of the observations in the sample. As we can see in the table, when the cut-off varies, the mean is relatively stable in the range of [0.56, 0.72]. Specifically, the mean of 'normal rainfall' defined by the IDM in the sample, which equals to 0.67, deviates by [-5, +9] when allowing the cut-off bounces down and up by 5% respectively. This difference is not particularly large, and hence we keep using the definition of 'normal rainfall' brought in by IDM. As our main interest is to study the coefficient of the interaction, to make the interpretation less difficult, we centre the two causal variables - rainfall and TSC - so that the coefficient on the interaction is the effect size at the mean.

To study the effect of rainfall in early life on human capital, we follow Shah and Steinberg (2017) and Kumar et al. (2016) by assigning in-utero 'normal rainfall' to children as a proxy for baseline stocks of human capital at birth, as denoted by θ_0 in Equation 4.3, using information of district of residence and year of birth of children.

In the period between 2000-2003 (i.e. when children in our sample were in utero), there are more than 70 per cent of the districts in our sample having variation of rainfall shocks, among which there are 13 per cent districts with only one year of normal rainfall, 30 per cent with two years of normal rainfall,

TABLE 4.1: Mean of normal rainfall in utero when the cut-off of the definition of normal rainfall varies

	Mean	Std. Dev.
Normal rainfall =1 if rain was above 70% of mean	0.72	0.45
Normal rainfall =1 if rain was above 71% of mean	0.71	0.45
Normal rainfall =1 if rain was above 72% of mean	0.70	0.46
Normal rainfall =1 if rain was above 73% of mean	0.69	0.46
Normal rainfall =1 if rain was above 74% of mean	0.64	0.48
Normal rainfall =1 if rain was above 75% of mean	0.67	0.47
Normal rainfall =1 if rain was above 76% of mean	0.64	0.48
Normal rainfall =1 if rain was above 77% of mean	0.62	0.49
Normal rainfall =1 if rain was above 78% of mean	0.61	0.49
Normal rainfall =1 if rain was above 79% of mean	0.59	0.49
Normal rainfall =1 if rain was above 80% of mean	0.56	0.50
Observations	3044	

Note: The sample is the one used in the main analysis. "Normal rainfall" =1 is for individuals whose in utero monsoon rainfall was above a certain percentage of the 34-year historical locality-specific mean. For example, if we follow the definition from India Meteorological Department (IDM), i.e. a moderate drought or severe drought occurs when the rainfall is below 75% of mean, then normal rainfall =1 if rain was above 75% of mean.

and 28 per cent with three years of normal rainfall. We also test explicitly for serial correlation of rainfall in these four years, considering that we want to make sure that we are not picking up the effects of shocks from any other years except the single shock in the year studied. We do not find that droughts this year can be predicted by droughts in the previous year or be correlated with drought in years ahead.[9]

4.4.4 IHDS survey data

The India Human Development Survey 2011/12 (IHDS-II) (Desai and Vanneman, 2015) is a nationally representative, multi-topic survey carried out by the University of Maryland. It studies 41,554 households in 1,503 villages, 971 urban neighbourhoods and 372 districts across India. There are three datasets being used in this study: individual data, household data and woman fertility history.[10]

Due to the reason that adverse rainfall mainly affects the rural and agriculture-dependent population and that the TSC programme is solely targeted at the rural population, we restrict our analytical sample to rural households. In the survey, according to the 2001 census, rural areas are defined as areas with a population less than 5,000 persons, or most male employment is agricultural.

[9]We use Pearson's Correlation test and find most of the correlations are insignificant, and the highest degree of correlation is below 0.3, which indicates that droughts between years are fairly independent.

[10]See data at http://www.icpsr.umich.edu/icpsrweb/DSDR/studies/36151. The data was downloaded on March 1, 2016.

4.4.4.1 Dependent variables

The IHDS has conducted tests for children aged 8 to 11 in rural districts, following the Annual Status of Education Report (ASER) tests, which measure children's educational achievements in India annually since 2005.[11] In this study, we use the second round of IHDS 2011/12. The reason why we do not use the first round of IHDS 2005/06 is that children in the first round are too old to be exposed to the TSC programme in their critical developmental period, while most children in the second round aged 8 to 11 are exposed to start-up of the TSC programme in their first year of life.

For eligible children aged 8 to 11, their reading, math, and writing skills are examined.[12] In the reading test, they are tested whether they could read letters, words, paragraphs or stories. Following Shah and Steinberg (2017), the total reading score is calculated by summarising the total ability they get in reading tests. For example, if children can read paragraphs, they get 3 scores, as this implicitly means that they can read letter, words, and paragraphs. The range of total reading score is from 0 to 4, as the reading ability increases monotonously with the score. A child is scored 0 if he/she is not able to read any of the letters, words, paragraphs, or stories; a child is scored 4 if he/she can read stories. Similarly, in the math test, children are given math tests in recognising numbers, solving subtraction, and division problems. The range of total math score is from 0 to 3. Finally, in the writing test, a child's writing ability is examined. The range of total writing score is from 0 to 2. If the child cannot write, he/she is given score 0; if he/she can write but making one or two mistakes, he/she is given 1 score; and if he/she can write without any mistake,

[11]See https://www.asercentre.org/. ASER has been conducted annually since 2005 on a large number of children in India. However, this is not public data, and other information besides test outcomes is less rich compared to the IHDS.

[12]See example of tests at http://www.asercentre.org/p/141.html. Multiple languages of tests are available.

he/she is given 2 scores. To ease interpretation of the results in the analysis, we standardise these scores of the tests.

We also exploit three binary variables from these three tests as additional indicators for cognitive achievement. They are the ability to read paragraphs, the ability to solve subtraction problems, and the ability to write. The probability of these events in our sample is between 0.45 and 0.73, validating our identification strategy using a linear model.[13]

4.4.4.2 Control variables

In the IHDS survey, a set of children, children's parents, households characteristics, and village traits are available. When studying the test outcomes, we control for gender, birth year, birth quarter, year of survey, age in months, language of tests, ever school, caste, parental education, material made of roof and wall, household composition variables, household size, household expenditure per capita, and village characteristics, such as hours of electricity available per day, percentage of usage of phones, material made of road, and seven programmes accessible that are related to health, nutrition childhood interventions.

4.4.5 Sample

Since we are looking at the impact of early life conditions at locality-level, it is vital that we limit the sample to the households who have not moved since the child was in utero, in order to capture the correct information of child's exposure to rainfall and TSC in fetal and first year of life.[14] Therefore, we

[13]It has be widely recognised that if the probabilities are moderate, like between 0.20 and 0.80, then the fitness of linear and logistic models are equal, and choice of a linear model is preferred so as to ease interpretation.

[14]There are only 90 observations which are dropped, due to migration after women's pregnancy.

generate a sample of 3,044 rural children born between 2001 and 2004 (aged 8 to 11) with test outcomes from 221 districts. We report the descriptive statistics for some key variables in Table 4.2.

As we can see from Figure 4.2, which maps the working sample of 221 districts which we are going to study, there are sufficient variations in TSC intensity (shown by the left-hand map) and rainfall (shown by the right-hand map) by district across time. The dark areas in the map depicting the increment of TSC intensity suggest that TSC is active in these areas, and the dark areas in the map depicting the number of positive rainfall mean that plenty rainfall occurs in these areas. Most importantly, the dark areas in both maps do not go along with each other in the same period of time. This gives us an impression that the implementation of TSC does not cooperate with the occurrence of droughts. This is also proven to be true when we test whether there is a systematic difference of rainfall between districts with or without the implementation of TSC by 2004. The t test, with t-statistic of 0.23, suggests that we can not reject the hypothesis of zero difference in rainfall in districts with TSC programme or not.[15]

In Figure 4.3, we depict the identification strategy graphically for our study, learning the effects of two inputs in early life - rainfall and TSC - and their potential interaction, using a child aged 11 in the survey year 2012 as an example. When the survey IHDS was carried out in 2012, the child born in 2001 and aged 11 was tested. We are interested at how rainfall shock in the 2000 when the child was in utero and the TSC programme in 2001 when the child was in the first year of life affect the child's test outcomes at the age of 11 (tested in 2012), and also whether the return to the TSC programme is boosted by the baseline of human capital stocks.

[15]There are 90 districts among the 221 districts in our sample without any TSC implementation throughout the four years between 2001 and 2004.

TABLE 4.2: Descriptive statistics for key variables

Variable	Mean	Std. Dev.
TSC latrines intensity in first year of life	0.16	0.49
Normal rainfall in utero (dv)	0.63	0.48
Reading total score (standardised)	0.00	1.00
Can read paragraphs (dv)	0.54	0.50
Math total score (standardised)	0.00	1.00
Can solve subtraction problems (dv)	0.45	0.50
Writing total score (standardised)	0.00	1.00
Can write (dv)	0.73	0.44
Male (dv)	0.53	0.50
Age in months	113.76	12.38
Forward castes	0.21	0.40
Backward castes	0.40	0.49
Currently at school (dv)	0.99	0.70
Mother: no education (dv)	0.44	0.50
Father: no education (dv)	0.23	0.42
Roof material: made of pucca (dv)	0.59	0.49
Wall material: made of pucca (dv)	0.64	0.48
No. of male adults	1.46	0.79
No. of female adults	1.53	0.78
No. of male children	1.53	0.96
No. of female children	1.41	1.12
No. of female teenagers	0.25	0.55
No. of male teenagers	0.17	0.42
No. of persons	6.35	2.33
Household expenditure per capita	9.61	0.57
Village: hours of electricity available per day	12.97	6.67
Village: percentage of usage of phones	84.64	17.27
Village: katcha road (dv)	0.11	0.31
Village: access to immunisation (dv)	0.94	0.24
Village: access to health checkups (dv)	0.85	0.36
Village: access to food supplement, nutrition (dv)	0.92	0.28
Village: access to growth monitoring (dv)	0.85	0.36
Village: access to early childhood interventions (dv)	0.86	0.35
Village: access to adolescent girls programme (dv)	0.49	0.50
Village: no. of immunisation campaigns	5.10	5.83
Observations	3044	

Note: "*Dv*" is denoted for dummy variable. "Normal rainfall" =1 is for individuals whose in *utero* monsoon rainfall was above 75% of the 44-year historical locality-specific mean, as the India Meteorological Department (IDM) defines a moderate drought or severe drought occurs when the rainfall is below 75% of mean. "TSC latrines intensity" is the number of TSC latrines per 100 capita.

FIGURE 4.2: Variation in TSC and rainfall between 2001-2004

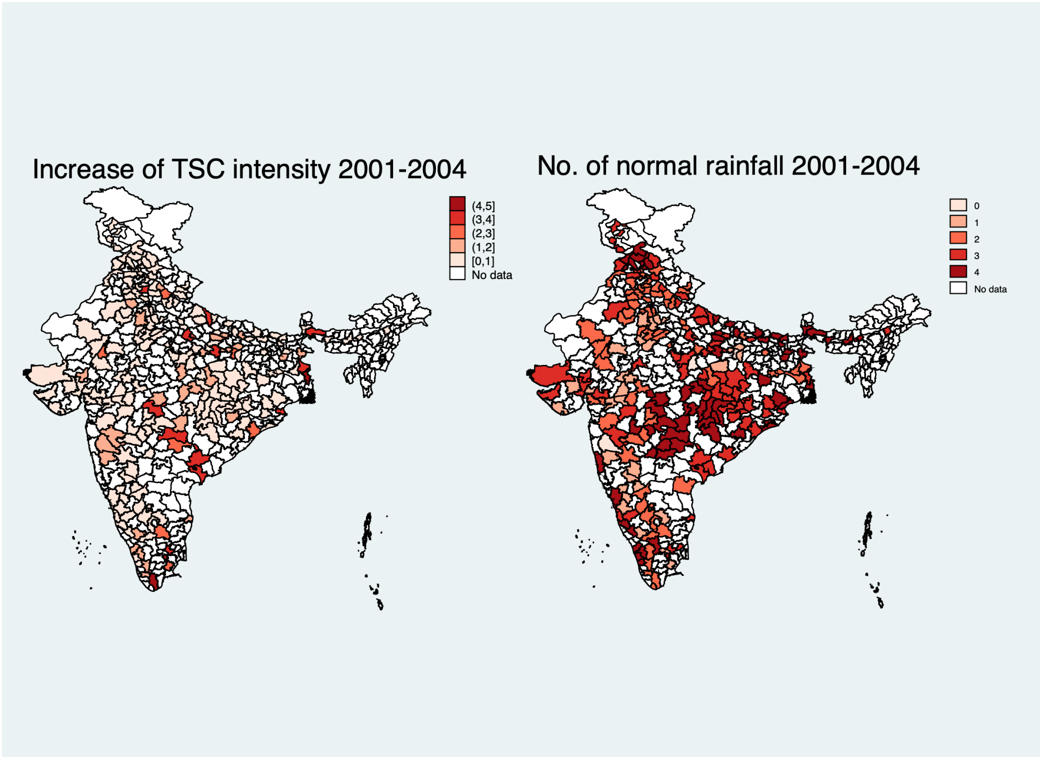
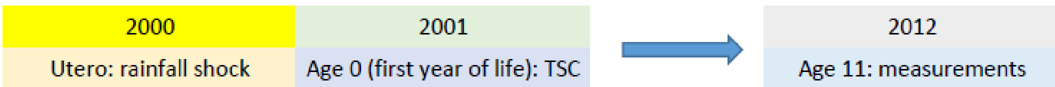


FIGURE 4.3: An illustration of the design of study



4.5 Empirical strategy

An important part of our empirical strategy lies in using the in utero rainfall shocks to capture an exogenous variation in the early life endowment. Shah and Steinberg (2017) have established that in rural India rainfall fluctuations affect wages and crop yield. As higher income will increase food consumption and nutritional input for fetus or infant, the authors further find that positive rainfall in early life will lead to better cognitive outcomes of rural children in the long term. We hence follow their strategy by using the quasi-random district-level rainfall in rural India as an exogenous shock to early life wages and therefore, biological endowments.

Secondly, we exploit the variation of TSC programme in the first year of life as an investment shock. Using difference-in-differences strategy, Spears and Lamba (2016) finds that TSC intensity at birth could increase child test outcomes and shows that the gradient of TSC-cognition is the steepest for exposure to the TSC in the first year of life and flat in other early-life years. Therefore, following this study, we use TSC intensity as an environmental investment shock, which also depends on birth year and residence of district.

Therefore, our basis empirical specification includes the rainfall shock, TSC intensity, and their interaction, same as the identification strategy in Adhvaryu et al. (2018), Duque et al. (2018), Gunnsteinsson et al. (2018), Johnson and Jackson (2017), Malamud et al. (2016), and Rossin-Slater and Wüst (2016). To validate a causal evaluation, we need to ensure that these variables induce exogenous variation in endowment and investments. Therefore it is essential that we check whether observables of children are systematically varying by TSC and rainfall by providing a test on the balance of covariates. Therefore, in Section 4.6.4.2, we report results of analysis regressing children, households and village characteristics which are described in Section 4.4.4.2 and included into the model as the control variables on the variables of interest: TSC intensity and rainfall. In summary, we find no significant difference in these characteristics across groups with high and low TSC intensity, as well as positive and negative rainfall shock groups.

We use a difference-in-differences (DID) estimation for our main specification, which is equivalent to adding district and birth year fixed effects to the main model. Therefore, separately for each outcome α, the specification will become the following:

$$H_{idt} = \beta_1 TSC_{dt} + \beta_2 R_{dt} + \beta_3 TSC_{dt} \times R_{dt} + \delta' X_{idt} + \nu_d + \gamma_t + \epsilon_{idt} \tag{4.5}$$

where H_{idt} represents measurements of capabilities for individual i in district d born in year t, i.e. test outcomes proxying for cognitive ability; TSC_{dt} is TSC latrines per 100 capita in district d in birth-year t; R_{dt} is a dummy variable depicting the rainfall in district d when children are in utero: it equals to 1 if there is a normal rainfall; 0 if there is a drought; $TSC_{dt} \times R_{dt}$ is the interaction between rainfall and TSC; X_{idt} are the controls using a series of children, parents, households, and village characteristics described in Section 4.4.4.2; ν_d is the district fixed-effects;γ_t captures the linear birth year trend; and finally ϵ_{idt} is an idiosyncratic error term. Standard errors are clustered at district level.

The argument that Equation 4.5 estimates the causal effects of rainfall, TSC, and their interaction is based on the assumption of parallel trends. It requires that on average, test scores would have evolved similarly in districts with or without TSC implementation by 2004 if TSC had not established. Albeit we cannot test it using a counterfactual case, we can instead look at whether the trends happened before TSC were similar across these two sets of districts. In Section 4.6.4.1, we test this using the test score of older cohorts who were born before the programme was initiated. We find that in these two sets of districts, test scores have had evolved similarly, despite the fact that the programme was started in districts or not.

Given the validation of the identification strategy, β_1 in Equation 4.5, denotes the main effect of a positive early-life nutritional input shock. While β_2 alone represents the effect of TSC for those who did not experience a positive nutritional input shock, $\beta_2 + \beta_3$ denotes that for those who experienced a positive shock. Our research interest mainly focuses on the interaction between rainfall and TSC. β_3 is the coefficient of interaction between rainfall and TSC at adjacent stages, and it shows the differential effect of TSC between high-endowed and low-endowed individuals. If it is positive, it implies *'dynamic complementarities'* and suggests that TSC had a larger effect for high-endowed children than low-endowed children; if it is negative, it implies *'dynamic substitutabilities'* and indicates that TSC can lead to a remediation for the negative effect of an early life adverse event.

4.6 Results

In this section, we start by discussing the results found using rainfall shocks defined above to examine whether early-life rainfall has an impact on test outcomes in our study. Then we report the results from the estimation of the impact of TSC using DID strategy discussed in the last section. Next, we report and discuss estimation results when both rainfall shocks and TSC are included in the model, and when their interaction is added afterwards. Finally, we report a number of checks to address concerns about the existence of a parallel trend and the balance of covariates.

4.6.1 Effects of rainfall on test outcomes

Table 4.3 presents the main estimates of the effect of in-utero rainfall on test scores using the cohorts of 2001-2004 who are aged 8-11, during which TSC was initiated in the early stage. In the first two columns, we examine the effect of rainfall on reading total scores and the ability to read paragraphs. The coefficient on normal rain in the first column is 0.076 with a 90% confidence interval of [0.004, 0.147], which implies that for an in-utero exposure to normal rainfall, children could score on average 7.6% of one standard deviation higher. However, the confidence interval is relatively large. The coefficient of the binary outcome 'can read paragraph', shown in column 2, is 0.039 with a 90% confidence interval of [0.004, 0.074]. This indicates that children are on average 4% more likely to be able to read paragraph if they have experienced

TABLE 4.3: Effects of rainfall on test outcomes

Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)
	Reading total score	Can read paragraphs (dv)	Math total score	Can do subtraction (dv)	Writing total score	Can write (dv)
Normal rainfall	0.076*	0.039*	-0.029	-0.003	0.064	0.028
	(0.043)	(0.021)	(0.041)	(0.019)	(0.044)	(0.019)
Observations	3044	3044	3033	3036	3024	3026
Fixed effects	District, birth year					
Birth year	2001-2004					
Ages	8-11					

Note: Standard errors clustered at districts are reported in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All models apply a district and birth year fixed-effects model. As in the following analysis to examine the impact of the TSC focusing on the first four years of the program, i.e. from 2001 to 2004, the sample is selected to be born among these four years. "Normal rainfall" =1 is for individuals whose in utero monsoon rainfall was above 75% of the 34-year historical locality-specific mean, as the India Meteorological Department (IDM) defines a moderate drought or severe drought occurs when the rainfall is below 75% of mean. All specifications include ever school, gender, birth year, birth quarter, year of survey, age in months, language of tests, caste, parental education, material made of roof and wall, household composition variables, household size, household expenditure per capita, and village characteristics (such as percentage of use of phones, hours of electricity available per day material made of road, and seven programmes related to health, nutrition and childhood development).

a normal rainfall year in utero, compare to ones experienced a drought year. In the sample, there are 54% of the children capable of reading paragraphs. The results are consistent with that of Shah and Steinberg (2017), suggesting that higher wages in early life could be translated into higher human capital investments and better educational achievement in later life.

However, no effect is found when using math test scores and writing test scores, nor using other binary outcomes, such as can solve subtraction problems and can write. Nonetheless, as shown in column 5 and 6, the standard errors of coefficients in the model using 'writing total score' and 'can write' are relatively small compare to the math results, resulting in the 90% confidence intervals of [-0.009, 0.137] and [-0.004, 0.060]. We reckon that the reasons why rainfall cannot predict writing test outcomes might be that the variation in these outcomes are relatively smaller than the ones of reading tests, or that variation in rainfall is not sufficient in explaining the variation in outcomes using the sample of cohorts of 2001-2004. Therefore, we expand the sample including cohorts born in 1999-2000 to increase the variation in rainfall and outcomes. We find that the effects on writing total score and ability to write are both significant at 95% significant level in the sample of cohorts born in 1999-2004. In Appendix Table E1, the coefficients suggest that in-utero exposure to normal rainfall would increase the writing total scores by 8.3% of one standard deviation, with a 90% confidence interval of [0.016, 0.151] and cause the probability of being able to write to increase by 3.6%, with a 90% confidence interval of [0.008, 0.065]. One should notice that although the magnitudes of the point estimates are large, these intervals are relatively wide. The effect size on the reading test and the ability of reading paragraph is relatively the same as the one found in the main results.

As empirical evidence suggests that effects of rainfall shocks on human capital are mainly found in utero or the first year of life, we also test whether rainfall fluctuations at birth has a similar impact on test outcomes. However, in Appendix Table E2, it is shown that there is no significant effect of rainfall in the first year of life. These results resonate with these two recent studies: Krutikova and Lilleør (2015) find that in-utero exposure to rainfall deviation is having an important impact on child's personal traits in adulthood using sibling-fixed effects in Tanzania, while no effect is found in the first year of life; Rosales (2014) also find similar results in that effect of negative rainfall shocks on children's health and cognitive outcomes is only found in utero but not the first year of life in Ecuador. Although Shah and Steinberg (2017) have found significant effect during in utero and the first three years of life, it might be the result of a large sample of more than 2 million, which is much bigger than our sample.

4.6.2 Effects of TSC on test outcomes

In this section, we report the effects of TSC on test outcomes. Using the same model as above but replacing rainfall with TSC, we find that TSC has a positive and significant impact on reading total scores and math total scores. As shown in Table 4.4, if there is an additional TSC latrine per 100 capita, the child could score 7.1% of one standard deviation higher in reading test on average, with a 90% confidence interval of [0.008, 0.134]; and 8.7% of one standard deviation higher in math test on average, with a 90% confidence interval of [0.021, 0.152]. These are substantial magnitudes, especially the standard error of coefficient for math total score is comparably small to the effect size. These results agree with the ones found by Spears and Lamba (2016).

As a robustness check, we also report the results when calibrating TSC intensity measured by a varying population across the birth years. Specifically, we replace the denominator in Equation 4.4, an unvarying population from

TABLE 4.4: Effects of TSC on test outcomes

Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)
	Reading total score	Can read paragraphs (dv)	Math total score	Can do subtraction (dv)	Writing total score	Can write (dv)
TSC latrines intensity	0.071*	0.029	0.087**	0.029	0.010	0.010
	(0.038)	(0.020)	(0.040)	(0.022)	(0.046)	(0.019)
Observations	3044	3044	3033	3036	3024	3026
Fixed effects	District, birth year					
Birth year	2001-2004					
Ages	8-11					

Note: Standard errors clustered at districts are reported in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All models apply a district and birth year fixed-effects model. To examine the impact of the TSC, we focus on the first four years of the program, i.e. from 2001 to 2004. Hence the sample is the cohorts born in these four years. "TSC latrines intensity" is the number of TSC latrines per 100 capita. All specifications include ever school, gender, birth year, birth quarter, year of survey, age in months, language of tests, caste, parental education, material made of roof and wall, household composition variables, household size, household expenditure per capita, and village characteristics (such as percentage of use of phones, hours of electricity available per day material made of road, and seven programmes related to health, nutrition and childhood development).

census 2001 with the corresponding population in birth year, which is calculated using the district-level growth rate computed by using census 2001 and 2011.[16] As shown in Appendix Table E3, the effects are consistently similar in size and significant level with the results reported above. Thus, in the following analysis, we continue to use the TSC intensity measured by population in 2001.

We also explore the effect by the timing of TSC exposure for mechanism check. We use the same test score data from IHDS and TSC administrative data as in the main analysis. Also, we keep the same cohorts studied as these are the ones being administered in tests in the survey. By using the same identification strategy, we try to study the potential effect of exposure to TSC in other years of life besides the birth year (age 0). Shown in Appendix Table E4, exposures in the second year (age 1) and third year (age 2) of life are not statistically significantly associated with cognitive outcomes measured by three tests. This is consistent with Spears and Lamba (2016)'s finding on mechanism check that the sanitation-cognition gradient is steepest in the first year of life. Moreover, it lifts concerns that our findings of the effect of TSC on learning outcomes are merely reflecting spurious correlations between sanitation trends and educational or socioeconomic trends, since the effect is only found when there is a proper matching of time of exposure.

[16]The denominator, an unvarying population, has been replaced by a varying population. Thus the TSC intensity is calculated by the following,

$$TSC_{dt} = \frac{\sum_{2001}^{t} \text{TSC latrines counts}_{dk}}{\text{rural population}_{dt}} \times 100$$

The annual population growth rate between 2001 and 2011 is the deviation between natural log of two years' population, divided by 10, as there are ten years in between.

4.6.3 Effects of rainfall and TSC on test outcomes

Having established that rainfall fluctuations and TSC are both relevant to early childhood development, we move on to our main analysis, reporting the results of the estimation using these two sources of shocks in early life and their interaction. Before discussing the results of Equation 4.5, the model where interactions are included, we first report the results of regressions including only the main effects of rainfall and TSC, shown in Panel A of Table 4.5.

The results in column 1 show that when a child experiences a normal rainfall in utero (i.e. no drought), the reading total score is likely to increase by 7.3% of one standard deviation; increasing one latrine per 100 capita leads individuals to raise reading score by 6.8% of one standard deviation. These coefficients are both significant at 90% confidence level, with standard error of 0.038 for the coefficient of TSC and 0.043 for the coefficient of rainfall. Moreover, the fact that the size of these effects are very close to the ones regressing on each shock solely, reported in Table 4.3 and 4.4, indicates that these two variables are independent, and therefore elevates concerns of dubious interaction effect in the following analysis when we add interactions into the model.

In column 2, we see that normal rainfall has positive impacts on child's ability in reading paragraphs. Individuals who did not experience adverse rainfall are 3.8% more likely to be able to read paragraphs, with a 90% confidence interval of [0.003, 0.073]. However, the coefficient of TSC is 0.028 with a 90% confidence interval of [-0.005, 0.060].

As shown in column 3-6, rainfall can not predict math and writing test outcomes, with insignificant coefficients, the same as the results found in Table 4.3.[17] Similar and significant coefficients are found for TSC as in Table 4.4. An additional TSC latrine leads to an average 8.8% of one standard deviation

[17]However, we are reminded from Appendix Table E1 that the writing tests are good indicators if there is sufficient variation.

TABLE 4.5: Effects of rainfall and TSC on test outcomes

Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)
	Reading total score	Can read paragraphs (dv)	Math total score	Can do subtraction (dv)	Writing total score	Can write (dv)
Panel A: Main effects only						
TSC latrines intensity	0.068*	0.028	0.088**	0.029	0.007	0.009
	(0.038)	(0.020)	(0.040)	(0.022)	(0.046)	(0.019)
Normal rainfall	0.073*	0.038*	-0.032	-0.004	0.064	0.028
	(0.043)	(0.021)	(0.041)	(0.019)	(0.044)	(0.019)
Panel B: Main effects and interactions						
TSC latrines intensity	0.049	0.015	0.078*	0.028	-0.011	0.004
	(0.039)	(0.022)	(0.041)	(0.022)	(0.044)	(0.020)
Normal rainfall	0.076*	0.039*	-0.030	-0.004	0.067	0.028
	(0.043)	(0.021)	(0.041)	(0.019)	(0.045)	(0.019)
Normal rainfall $\times$ TSC	0.100	0.065	0.053	0.008	0.096	0.025
	(0.074)	(0.041)	(0.079)	(0.045)	(0.087)	(0.042)
Observations	3044	3044	3033	3036	3024	3026
Fixed effects	District, birth year					

Note: Standard errors clustered at districts are reported in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All models apply a district and birth year fixed-effects model. To examine the impact of the TSC, we focus on the first four years of the program, i.e. from 2001 to 2004. Hence the sample is the cohorts born in these four years. "Normal rainfall" =1 is for individuals whose in utero monsoon rainfall was above 75% of the 34-year historical locality-specific mean, as the India Meteorological Department (IDM) defines a moderate drought or severe drought occurs when the rainfall is below 75% of mean. "TSC latrines intensity" is the number of TSC latrines per 100 capita. All specifications include ever school, gender, birth year, birth quarter, year of survey, age in months, language of tests, caste, parental education, material made of roof and wall, household composition variables, household size, household expenditure per capita, and village characteristics (such as percentage of use of phones, hours of electricity available per day material made of road, and seven programmes related to health, nutrition and childhood development).

increase in math total score, significant at 95% confidence level, as shown in column 3. The 90% confidence interval is [0.022, 0.154]. In column 5 and 6, the results suggest that the coefficients of the TSC programme are not significant at 10%.

In Panel B of Table 4.5, we report results of Equation 4.5, which sheds light on the main interest of this paper. Heterogeneous results are displayed when interaction terms are added into the specifications, and we find that the coefficients of the interaction terms are all positive but non-significant. These results are caused by the large standard errors of coefficients. Specifically, the 90% confidence intervals of the effect size in column 1 and 2 are [-0.022, 0.222] and [-0.003, 0.134] respectively, where the point estimate is 0.100 and 0.065. Although the results of interaction terms cannot provide us a confident conclusion of complementarity in investments, there is indication of the existence of complementarity, hinted by the the changing effect size of TSC in column 3. The fact that the coefficient in Panel B (=0.078) is smaller than the main effect in Panel A (=0.088) suggests that TSC exposure has a larger impact on children with a higher endowment as the heterogeneous effect 'took away' some effect of the overall TSC effect.

In Table 4.6, we report results of the estimation using the same identification strategy, but with a set of combined test scores as outcome variables, representing the overall cognitive ability across subjects. These scores are standardised. In column 1, we report the results of specification with total scores of three tests as an overall cognition outcome in reading, writing, and math. In column 2, the outcome variable is the summation of the test scores from reading and writing tests, as both of these could reflect the facet of language skill of human capital, such as orthographic and phonological skills (Mäki et al., 2001). For example, the deficiency in these skills could affect reading and writing simultaneously (Stage and Wagner, 1992). Also, the quality of writing can be improved by the ability to read as children can revise their writing by evaluative reading, which can reduce their writing mistakes (Hayes et al., 1987; Hayes, 2000). Therefore, we proxy the development of language skill by the summation of reading and writing tests, which might hopefully reflect the level of human cognition in language more precisely.

In Panel A, shown in column 1 and 3, an additional TSC latrine can increase child's total scores by 6.7% of one standard deviation of total scores, with a standard error of 0.038, and increase total read and math score by 8.3% of one standard deviation, with a standard error of 0.037. Shown in column 2, normal rainfall would increase total read and write score by 7.5% of one standard deviation, with a standard error of 0.042. In Panel B, we find similar results as in Table table:inter. The coefficients of the interaction terms are positive yet non-significant. Due to the large standard errors, the 90% confidence intervals of the interaction terms are [-0.033, 0.228], [-0.016, 0.231], [-0.038, 0.215] and [-0.064, 0.226]. However, in column 3, we could also find that the coefficients of TSC drop from 0.083 to 0.066 after adding the interaction term, implying a potential existence of complementarity.

These results add evidence to the plausible results we find in Table 4.5 that the interaction effects are positive, and might support our hypothesis that there is an endowment gradient in the treatment effect. However, due to the large standard errors of the interaction terms, we lack of confidence to conclude that the return to investment is heterogeneous across levels of endowments and that it is the high-endowed children who respond more strongly to the exposure of better disease environment in the first year of their life. Yet, this is not surprising since one similar study exploits the access to abortion and to better school in Romania and finds no significant interaction effect (Malamud et al., 2016).

TABLE 4.6: Effects of rainfall and TSC on combined scores

Dependent variable	(1)	(2)	(3)	(4)
	Total scores	Read & write	Read & math	Write & math
Panel A: Main effects only				
TSC latrines intensity	0.067*	0.049	0.083**	0.058
	(0.038)	(0.040)	(0.037)	(0.038)
Normal rainfall	0.042	0.075*	0.030	0.012
	(0.041)	(0.042)	(0.042)	(0.041)
Panel B: Main effects and interactions				
TSC latrines intensity	0.049	0.029	0.066*	0.043
	(0.038)	(0.039)	(0.038)	(0.039)
Normal rainfall	0.044	0.078*	0.033	0.014
	(0.041)	(0.042)	(0.042)	(0.041)
Normal rainfall $\times$ TSC	0.098	0.107	0.089	0.081
	(0.079)	(0.075)	(0.077)	(0.088)
Observations	3018	3024	3033	3018
Fixed effects	District, birth year			

Note: Standard errors clustered at districts are reported in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All models apply a district and birth year fixed-effects model. To examine the impact of the TSC, we focus on the first four years of the program, i.e. from 2001 to 2004. Hence the sample isthe cohorts born in these four years. "Normal rainfall" =1 is for individuals whose in utero monsoon rainfall was above 75% of the 34-year historical locality-specific mean, as the India Meteorological Department (IDM) defines a moderate drought or severe drought occurs when the rainfall is below 75% of mean. "TSC latrines intensity" is the number of TSC latrines per 100 capita. All specifications include ever school, gender, birth year, birth quarter, year of survey, age in months, language of tests, caste, parental education, material made of roof and wall, household composition variables, household size, household expenditure per capita, and village characteristics (such as percentage of use of phones, hours of electricity available per day material made of road, and seven programmes related to health, nutrition and childhood development).

4.6.4 Robustness checks

4.6.4.1 Parallel trends in test outcomes in older children

To validate the estimates of the effect of the TSC, we provide evidence for the existence of parallel trends in this section. The DID estimation relies on the assumption of parallel trends, which assumes that if TSC had not happened, the trends in the evolution of districts would have been similar regardless of the TSC implementation. Though no direct test of counterfactual assumption is possible, we provide evidence indicating that parallel trends exist before TSC implementation using cohorts born before 2001. As the IHDS data also administered tests to children born between 1999-2000, we take them as our sample, who would have been too old to be exposed to TSC as shown in the mechanism check of TSC in Section 4.6.2.

We want to check whether districts with TSC implementation are correlated with time trends so that our results might merely reflect district time trends. That is to say, if there is a pre-existing heterogeneous effect of time trends by TSC implementation in this sample, it would suggest that TSC implementation is correlated with pre-programme trends in test outcomes.

In Table 4.7, we report results of the following regression,

$$testscore_{idst} = \beta_1 NoTSC_{ds} + \beta_2 NoTSC_{ds} \times time_{it} + \delta' X_{idst} + \nu_s + \gamma_t + \epsilon_{idst} \tag{4.6}$$

where i indexes individual children, d districts, s states, and t age cohorts; $NoTSC_{ds}$ is a district-level time-invariant dummy variable, which equals to 1 if there is no TSC implementation by 2004 and 0 otherwise[18]; $time_{it}$ is the dummy variable capturing the time trend between 1999-2000, which equals to

[18]There are in total 147 districts in this sample, among which 54 are without TSC by 2004.

TABLE 4.7: Parallel trends: using cohorts born before TSC's implementation

Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Reading total scores	Math total scores	Writing total scores	Total scores	Read & write	Read & math	Write & math
Panel A: Birth year fixed-effects							
No TSC (dv)	0.081	-0.132	-0.503	-0.193	-0.195	-0.049	-0.344
	(0.379)	(0.402)	(0.393)	(0.419)	(0.435)	(0.415)	(0.414)
No TSC × time	-0.103	0.355	0.466	0.260	0.172	0.135	0.472
	(0.410)	(0.442)	(0.421)	(0.441)	(0.453)	(0.441)	(0.449)
Fixed effects	Birth year						
Panel B: State, birth year fixed-effects							
No TSC (dv)	0.043	-0.083	-0.492	-0.199	-0.239	-0.082	-0.290
	(0.367)	(0.330)	(0.392)	(0.367)	(0.416)	(0.345)	(0.351)
No TSC × time	-0.139	0.250	0.414	0.198	0.149	0.089	0.371
	(0.393)	(0.361)	(0.407)	(0.380)	(0.427)	(0.363)	(0.385)
Fixed effects	State, birth year						
Observations	435	433	431	430	431	433	430

Note: Standard errors clustered at state-level are reported in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. In Panel A, models apply a birth year fixed-effects model, and in Panel B, an inclusion of state FE. The sample is the older children who are born before TSC was implemented, i.e. 1999-2000, and who were administered in tests. "No TSC (dv)" is a district-level time-invariant dummy variable, which equals to 1 if there is no TSC implementation by 2004. There are in total 147 districts in this sample, among which 54 are without TSC by 2004. "Time" is the dummy variable capturing the time trend between 1999-2000, which equals to 1 if the child is born in 2000. All specifications include ever school, gender, birth year, birth quarter, year of survey, age in months, language of tests, caste, parental education, material made of roof and wall, household composition variables, household size, household expenditure per capita, and village characteristics (such as percentage of use of phones, hours of electricity available per day material made of road, and seven programmes related to health, nutrition and childhood development).

1 if the child is born in 2000 and 0 otherwise; $NoTSC_{ds} \times time_{it}$ is the within-district time trend; X_{idst} is the set of controls same as the main model; ν_s is the state fixed-effects; and γ_t is the birth year fixed-effects. Standard errors are clustered by state.

In Panel A, we report results of regressions using only birth year fixed-effects; in Panel B, state fixed-effects are included in the models. As we can see from both panels, across seven different measures of cognitive achievement, no district trends are found as the coefficients of the interaction of TSC implementation and time β_2 are all insignificant, using only year fixed-effects or state and year fixed-effects. This supplies evidence for the existence of the parallel trend that the TSC implementation was not correlated with preexisting trends in cognition development.

4.6.4.2 Balance of covariates

To ensure that our identification strategy for studying the effects of TSC and rainfall is valid, we provide analysis on whether TSC and rainfall exogenously impact on children cognitive outcomes by testing the association between TSC intensity, positive rainfall and other observed variables describing child characteristics and environment. Specifically, we test whether the covariates that are correlated with child outcomes can be predicted by TSC and rain. Since we do not have a randomised trial, in which a direct check on the balance of observables between treatment and control groups can be provided directly, we instead present results regressing these covariates on TSC and rain as an analogy. In Appendix Table E5, estimates from separate specifications regressing each control variable on TSC and rain are shown, conditional on district, birth year, year of survey, and state $\times$ birth year fixed-effects.

Across a total of 39 coefficients, only five are significant. The proportion of significant coefficients is what we expect to see. Most importantly, the size of these coefficients are small compared to the means. The results suggest that neither the TSC or rainfall could predict the observable characteristics of children and households. Particularly, children exposed to good TSC environment or good rainfall are not more likely to be enrolled at school, nor are they born in families with more children. Observable village traits are not correlated with TSC and rain indicating that the effect of TSC we find does not reflect trends in local interventions and programmes and that rainfall is exogenously distributed at locality level. We realise that this analysis could not fully relieve one's concern of the possibility of omitted factors which could bias the results, yet it is very unlikely as such an important variable is not captured and not correlated with the observed variables we examine.

4.7 Conclusion

In this paper, we study whether the effects of early life shocks on cognitive achievements are mitigated or amplified by later investments, or more specifically, whether the return to investments varies by level of endowments. We provide reduced-form estimates from an identification strategy relying on the exogenous variation in endowment and investment at individual-level generated by random variation in rainfall at district-level and a policy programme in constructing latrines, these both conditional on district and birth year fixed-effects.

We find positive effects of rainfall in utero and TSC latrines intensity at birth in producing cognitive outcomes in later childhood. Although we have found positive coefficients of the interaction terms, due to the large standard errors of the interaction terms, we could not be confident to conclude that there is a heterogeneous effect of investments such that high-endowed children benefit more from TSC than low-endowed children. This suggests that disadvantaged children could generate as much of the profit from the TSC programme as the advantaged children. These results suggest the absence of 'dynamic complementarities' in human capital formation, which is consistent with what Malamud et al. (2016) find.

However, it is not clear whether the disadvantaged children can still responsive to investments in later life as they age and whether the type of investments matter for the interactive effect (Duque et al., 2018). In future research, as noted by Almond et al. (2018), potential heterogeneous effects of treatments in the interval between early life and adulthood life could be further explored when more comprehensive administrative data is accessible.

Bibliography

Adhvaryu, Achyuta, Teresa Molina, Anant Nyshadham, and Jorge Tamayo (2018). *Helping children catch up: Early life shocks and the PROGRESA Experiment*. Working Paper w24848. National Bureau of Economic Research.

Aizer, Anna and Flávio Cunha (2012). *The production of human capital: Endowments, investments and fertility*. Working Paper 18429. National Bureau of Economic Research.

Aizer, Anna, Shari Eli, Joseph Ferrie, and Adriana Lleras-Muney (2016). "The long-run impact of cash transfers to poor families". *The American Economic Review* 106.4, pp. 935–971.

Almond, Douglas (2006). "Is the 1918 influenza pandemic over? Long-term effects of in utero influenza exposure in the post-1940 US population". *Journal of Political Economy* 114.4, pp. 672–712.

Almond, Douglas and Janet Currie (2011). "Killing me softly: The fetal origins hypothesis". *The Journal of Economic Perspectives* 25.3, pp. 153–172.

Almond, Douglas and Bhashkar Mazumder (2013). "Fetal origins and parental responses". *Annual Review of Economics* 5.1, pp. 37–56.

Almond, Douglas, Lena Edlund, and Mårten Palme (2009). "Chernobyl's subclinical legacy: Prenatal exposure to radioactive fallout and school outcomes in Sweden". *The Quarterly Journal of Economics* 124.4, pp. 1729–1772.

Almond, Douglas, Janet Currie, and Valentina Duque (2018). "Childhood circumstances and adult outcomes: Act II". *Journal of Economic Literature* 56.4, pp. 1360–1446.

Attri, SD and Ajit Tyagi (2010). "Climate profile of India". *Environment Monitoring and Research Center, India Meteorology Department: New Delhi, India*.

Auffhammer, Maximilian, Veerabhadran Ramanathan, and Jeffrey R Vincent (2012). "Climate change, the monsoon, and rice yield in India". *Climatic Change* 111.2, pp. 411–424.

Augsburg, Britta and Paul Andrés Rodríguez-Lesmes (2018). "Sanitation and child health in India". *World Development* 107, pp. 22–39.

Banerjee, Abhijit, Esther Duflo, Gilles Postel-Vinay, and Tim Watts (2010). "Long-run health impacts of income shocks: Wine and phylloxera in nineteenth-century France". *The Review of Economics and Statistics* 92.4, pp. 714–728.

Barker, DJP (1995). "Fetal origins of coronary heart disease". *British Medical Journal (BMJ)* 311.6998, pp. 171–174.

Barnard, Sharmani, Parimita Routray, Fiona Majorin, Rachel Peletz, Sophie Boisson, Antara Sinha, and Thomas Clasen (2013). "Impact of Indian Total Sanitation Campaign on latrine coverage and use: A cross-sectional study in Orissa three years following programme implementation". *PloS one* 8.8, e71438.

Checkley, William et al. (2008). "Multi-country analysis of the effects of diarrhoea on childhood stunting". *International Journal of Epidemiology* 37.4, pp. 816–830.

Chen, Yuyu and Li-An Zhou (2007). "The long-term health and economic consequences of the 1959–1961 famine in China". *Journal of Health Economics* 26.4, pp. 659–681.

Clasen, Thomas et al. (2014). "Effectiveness of a rural sanitation programme on diarrhoea, soil-transmitted helminth infection, and child malnutrition in Odisha, India: A cluster-randomised trial". *The Lancet Global Health* 2.11, e645–e653.

Colmer, Jonathan (2018). *Weather, labor reallocation and industrial production: Evidence from India*. LSE Research Online Documents on Economics. London School of Economics and Political Science, LSE Library.

Conti, Gabriella and James J Heckman (2014). *Economics of child well-being*. Springer.

Cumming, Oliver and Sandy Cairncross (2016). "Can water, sanitation and hygiene help eliminate stunting? Current evidence and policy implications". *Maternal & Child Nutrition* 12, pp. 91–105.

Cunha, Flavio and James Heckman (2007). "The technology of skill formation". *American Economic Review* 97.2, pp. 31–47.

Cunha, Flavio, James J Heckman, Lance Lochner, and Dimitriy V Masterov (2006). "Interpreting the evidence on life cycle skill formation". *Handbook of the Economics of Education* 1, pp. 697–812.

Cunha, Flavio, James J Heckman, and Susanne M Schennach (2010). "Estimating the technology of cognitive and noncognitive skill formation". *Econometrica* 78.3, pp. 883–931.

Currie, J et al. (2009). "Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development." *Journal of Economic Literature* 47.1, pp. 87–117.

Currie, Janet and Douglas Almond (2011). "Human capital development before age five". *Handbook of Labor Economics*. Vol. 4. Elsevier, pp. 1315–1486.

Currie, Janet and Tom Vogl (2013). "Early-life health and adult circumstance in developing countries". *Annual Review of Economics* 5.1, pp. 1–36.

Dercon, Stefan and Catherine Porter (2014). "Live aid revisited: Long-term impacts of the 1984 Ethiopian famine on children". *Journal of the European Economic Association* 12.4, pp. 927–948.

Desai, Sonalde and Reeve Vanneman (2015). "India Human Development Survey-II (IHDS-II), 2011-12. ICPSR36151-v2". *Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]* 31.

Doyle, Orla, Colm P Harmon, James J Heckman, and Richard E Tremblay (2009). “Investing in early human development: Timing and economic efficiency”. *Economics & Human Biology* 7.1, pp. 1–6.

Duque, Valentina, Maria Rosales-Rueda, and Fabio Sanchez (2018). *How do early-life shocks interact with subsequent human-capital investments? Evidence from administrative data*. Tech. rep. Mimeo, University of Michigan.

Grantham-McGregor, Sally M, Ernesto Pollitt, Theodore D Wachs, Samuel J Meisels, and Keith G Scott (1999). “Summary of the scientific evidence on the nature and determinants of child development and their implications for programmatic interventions with young children”. *Food and Nutrition Bulletin* 20.1, pp. 4–6.

Guerrant, David I, Sean R Moore, Aldo A Lima, Peter D Patrick, John B Schorling, and Richard L Guerrant (1999). “Association of early childhood diarrhea and cryptosporidiosis with impaired physical fitness and cognitive function four-seven years later in a poor urban community in northeast Brazil.” *The American Journal of Tropical Medicine and Hygiene* 61.5, pp. 707–713.

Guiteras, Raymond (2009). “The impact of climate change on Indian agriculture”. *Manuscript, Department of Economics, University of Maryland, College Park, Maryland*.

Gunnsteinsson, Snaebjorn, Achyuta Adhvaryu, Parul Christian, Alain Labrique, Jonathan Sugimoto, Abu Ahmed Shamim, and Keith P West Jr (2018). *Protecting Infants from Natural Disasters: The Case of Vitamin A Supplementation and a Tornado in Bangladesh*. Tech. rep. Working paper.

Hayes, John R (2000). “A new framework for understanding cognition and affect in writing”. *Perspectives on Writing: Research, Theory, and Practice* 6.

Hayes, John R, Linda Flower, Karen A Schriver, James Stratman, Linda Carey, et al. (1987). "Cognitive processes in revision". *Advances in Applied Psycholinguistics* 2, pp. 176–240.

Heckman, James J (2006). "Skill formation and the economics of investing in disadvantaged children". *Science* 312.5782, pp. 1900–1902.

— (2007). "The economics, technology, and neuroscience of human capability formation". *Proceedings of the National Academy of Sciences of the United States of America*, pp. 13250–13255.

Heckman, James J and Stefano Mosso (2014). "The economics of human development and social mobility". *Annu. Rev. Econ.* 6.1, pp. 689–733.

Humphrey, Jean H (2009). "Child undernutrition, tropical enteropathy, toilets, and handwashing". *The Lancet* 374.9694, pp. 1032–1035.

Jacoby, Hanan G and Emmanuel Skoufias (1997). "Risk, financial markets, and human capital in a developing country". *The Review of Economic Studies* 64.3, pp. 311–335.

Jayachandran, Seema (2006). "Selling labor low: Wage responses to productivity shocks in developing countries". *Journal of Political Economy* 114.3, pp. 538–575.

Jensen, Robert (2000). "Agricultural volatility and investments in children". *American Economic Review* 90.2, pp. 399–404.

Johnson, Rucker C and C. Kirabo Jackson (2017). *Reducing inequality through dynamic complementarity: Evidence from Head Start and public school spending*. Working Paper 23489. National Bureau of Economic Research.

Krutikova, Sonya and Helene Bie Lilleør (2015). *Fetal origins of personality: Effects of early life circumstances on adult personality traits*. Working Papers No. 2015-03. Centre for the Study of African Economies, University of Oxford.

Kumar, Santosh and Sebastian Vollmer (2013). “Does access to improved sanitation reduce childhood diarrhea in rural India?” *Health Economics* 22.4, pp. 410–427.

Kumar, Santosh, Ramona Molitor, and Sebastian Vollmer (2016). “Drought and early child health in rural India”. *Population and Development Review* 42.1, pp. 53–68.

Links, Knowledge (2008). “Feeling the pulse: A Study of the Total Sanitation Campaign in five states”. *New Delhi: WaterAid, November*.

Maccini, Sharon and Dean Yang (2009). “Under the weather: Health, schooling, and economic consequences of early-life rainfall”. *American Economic Review* 99.3, pp. 1006–26.

Mäki, Hanna S, Marinus JM Voeten, Marja MS Vauras, and Elisa H Poskiparta (2001). “Predicting writing skill development with word recognition and preschool readiness skills”. *Reading and Writing* 14.7-8, pp. 643–672.

Malamud, Ofer, Cristian Pop-Eleches, and Miguel Urquiola (2016). *Interactions between family and school environments: Evidence on dynamic complementarities?* Working Paper 22112. National Bureau of Economic Research.

Mbuya, Mduduzi NN and Jean H Humphrey (2016). “Preventing environmental enteric dysfunction through improved water, sanitation and hygiene: An opportunity for stunting reduction in developing countries”. *Maternal & Child Nutrition* 12, pp. 106–120.

Mondal, Dinesh, Juliana Minak, Masud Alam, Yue Liu, Jing Dai, Poonum Korpe, Lei Liu, Rashidul Haque, and William A Petri Jr (2011). “Contribution of enteric infection, altered intestinal barrier function, and maternal malnutrition to infant malnutrition in Bangladesh”. *Clinical Infectious Diseases* 54.2, pp. 185–192.

Patil, Sumeet R, Benjamin F Arnold, Alicia L Salvatore, Bertha Briceno, Sandipan Ganguly, John M Colford Jr, and Paul J Gertler (2014). “The effect of India's Total Sanitation Campaign on defecation behaviors and child health in rural Madhya Pradesh: A cluster randomized controlled trial". *PLoS Medicine* 11.8, e1001709.

Paxson, Christina H (1992). "Using weather variability to estimate the response of savings to transitory income in Thailand". *The American Economic Review*, pp. 15–33.

Petri, William A, Mark Miller, Henry J Binder, Myron M Levine, Rebecca Dillingham, and Richard L Guerrant (2008). "Enteric infections, diarrhea, and their impact on function and development". *The Journal of Clinical Investigation* 118.4, pp. 1277–1290.

Pickering, Amy J, Habiba Djebbari, Carolina Lopez, Massa Coulibaly, and Maria Laura Alzua (2015). "Effect of a community-led sanitation intervention on child diarrhoea and child growth in rural Mali: A cluster-randomised controlled trial". *The Lancet Global Health* 3.11, e701–e711.

Rocha, Rudi and Rodrigo R Soares (2015). "Water scarcity and birth outcomes in the Brazilian semiarid". *Journal of Development Economics* 112, pp. 72–91.

Rosales, María Fernanda (2014). *Impact of early life shocks on human capital formation*. Working Papers No. IDB-WP-503. Inter-American Development Bank.

Rosenzweig, Mark R and Hans P Binswanger (1992). *Wealth, weather risk, and the composition and profitability of agricultural investments*. Vol. 1055. World Bank Publications.

Rossin-Slater, Maya and Miriam Wüst (2016). *What is the added value of preschool for poor children? Long-term and intergenerational impacts and interactions with an infant health intervention*. Working Paper 22700. National Bureau of Economic Research.

Shah, Manisha and Bryce Millett Steinberg (2017). "Drought of opportunities: contemporaneous and long-term impacts of rainfall shocks on human capital". *Journal of Political Economy* 125.2, pp. 527–561.

Spears, Dean (2012a). "Effects of rural sanitation on infant mortality and human capital: Evidence from India's Total Sanitation Campaign". *Washington, DC: Princeton University*.

— (2012b). "Height and cognitive achievement among Indian children". *Economics & Human Biology* 10.2, pp. 210–219.

Spears, Dean and Sneha Lamba (2016). "Effects of early-life exposure to sanitation on childhood cognitive skills: Evidence from India's Total Sanitation Campaign". *Journal of Human Resources* 51.2, pp. 298–327.

Stage, Scott A and Richard K Wagner (1992). "Development of young children's phonological and orthographic knowledge as revealed by their spellings." *Developmental Psychology* 28.2, p. 287.

Thomas, Duncan, Kathleen Beegle, Elizabeth Frankenberg, Bondan Sikoki, John Strauss, and Graciela Teruel (2004). "Education in a crisis". *Journal of Development Economics* 74.1, pp. 53–85.

UNICEF and WHO (2014). *Progress on sanitation and drinking water: 2014 update*. World Health Organization.

Vyas, Sangita, Phyrum Kov, Susanna Smets, and Dean Spears (2016). "Disease externalities and net nutrition: Evidence from changes in sanitation and child height in Cambodia, 2005–2010". *Economics & Human Biology* 23, pp. 235–245.

Appendix

TABLE E1: Effects of rainfall on test outcomes using cohorts born between 1999-2004

Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)
	Reading total score	Can read paragraphs (dv)	Math total score	Can do subtraction (dv)	Writing total score	Can write (dv)
Normal rainfall 0.085**	0.043**	-0.000	0.008	0.083**	0.036**	
	(0.039)	(0.019)	(0.037)	(0.017)	(0.041)	(0.017)
Observations	3508	3673	3495	3664	3484	3652
Fixed effects	District, birth year					
Birth year	1999-2004					
Ages	8-13					

Note: Standard errors clustered at districts are reported in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All models apply a district and birth year fixed-effects model. The sample is cohorts born in the period between 1999-2004. "Normal rainfall" =1 is for individuals whose in utero monsoon rainfall was above 75% of the 34-year historical locality-specific mean, as the India Meteorological Department (IDM) defines a moderate drought or severe drought occurs when the rainfall is below 75% of mean. All specifications include ever school, gender, birth year, birth quarter, year of survey, age in months, language of tests, caste, parental education, material made of roof and wall, household composition variables, household size, household expenditure per capita, and village characteristics (such as percentage of use of phones, hours of electricity available per day material made of road, and seven programmes related to health, nutrition and childhood development).

TABLE E2: Effects of rainfall on test outcomes using rainfall in the first year of life

Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)
	Reading total score	Can read paragraphs (dv)	Math total score	Can do subtraction (dv)	Writing total score	Can write (dv)
Normal rainfall -0.028	-0.011	-0.027	-0.042	-0.060	-0.014	
	(0.053)	(0.026)	(0.054)	(0.027)	(0.050)	(0.022)
Observations	3044	3044	3033	3036	3024	3026
Fixed effects	District, birth year					
Birth year	2001-2004					
Ages	8-11					

Note: Standard errors clustered at districts are reported in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All models apply a district and birth year fixed-effects model. The sample is the same as used in Table 4.3. "Normal rainfall" =1 is for individuals whose first year of life monsoon rainfall was above 75% of the 34-year historical locality-specific mean, as the India Meteorological Department (IDM) defines a moderate drought or severe drought occurs when the rainfall is below 75% of mean. All specifications include ever school, gender, birth year, birth quarter, year of survey, age in months, language of tests, caste, parental education, material made of roof and wall, household composition variables, household size, household expenditure per capita, and village characteristics (such as percentage of use of phones, hours of electricity available per day material made of road, and seven programmes related to health, nutrition and childhood development).

TABLE E3: Effects of TSC on test outcomes using TSC measured by varying population

Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)
	Reading total score	Can read paragraphs (dv)	Math total score	Can do subtraction (dv)	Writing total score	Can write (dv)
TSC latrines intensity	0.072*	0.030	0.089**	0.030	0.012	0.011
	(0.038)	(0.020)	(0.039)	(0.022)	(0.046)	(0.019)
Observations	3044	3044	3033	3036	3024	3026
Fixed effects	District, birth year					
Birth year	2001-2004					
Ages	8-11					

Note: Standard errors clustered at districts are reported in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All models apply a district and birth year fixed-effects model. The sample is the same as used in Table 4.4. "TSC latrines intensity" is the number of TSC latrines per 100 capita, which district-level population varies across years. All specifications include ever school, gender, birth year, birth quarter, year of survey, age in months, language of tests, caste, parental education, material made of roof and wall, household composition variables, household size, household expenditure per capita, and village characteristics (such as percentage of use of phones, hours of electricity available per day material made of road, and seven programmes related to health, nutrition and childhood development).

TABLE E4: Sanitation-cognition gradient at the age of one and two

Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)
	Reading total score	Can read paragraphs (dv)	Math total score	Can do subtraction (dv)	Writing total score	Can write (dv)
Panel A: TSC at the age of one						
TSC latrine intensity (age 1)	0.000	-0.002	-0.002	-0.005	0.004	-0.001
	(0.027)	(0.013)	(0.030)	(0.016)	(0.029)	(0.012)
Panel B: TSC at the age of two						
TSC latrine intensity (age 2)	-0.002	-0.011	-0.007	-0.005	-0.009	0.004
	(0.013)	(0.007)	(0.013)	(0.007)	(0.012)	(0.006)
Observations	3044	3044	3033	3036	3024	3026
Fixed effects	District, birth year					
Birth year	2001-2004					
Ages	8-11					

Note: Standard errors clustered at districts are reported in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All models apply a district and birth year fixed-effects model. The sample is the same as used in Table 4.4. In panel A, the TSC administrative data for the independent variable is moved one year later so that the TSC exposure happens in age 1. In panel B, the TSC is moved two year later so that the TSC exposure happens in age 2. "TSC latrines intensity" is the number of TSC latrines per 100 capita. All specifications include ever school, gender, birth year, birth quarter, year of survey, age in months, language of tests, caste, parental education, material made of roof and wall, household composition variables, household size, household expenditure per capita, and village characteristics (such as percentage of use of phones, hours of electricity available per day material made of road, and seven programmes related to health, nutrition and childhood development).

TABLE E5: Balance of covariates: Rainfall shocks, TSC and control variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
	Male	Age in months	Never enrolled	Currently enrolled	Enrolled in past	Caste Brahmin	Caste Forward	Caste Backward	Caste SC	Caste ST	Caste others	Quarter first	Quarter second	Quarter third	Quarter forth	Mom no edu	Dad no edu	Roof pucca	Wall pucca	No. male adult
TSC	0.012	-0.093	0.003	-0.011	0.008	-0.009	-0.002	-0.019	0.003	0.027	0.000	-0.016	-0.015	0.040	-0.009	-0.024	-0.024	0.122***	0.029	-0.083*
	(0.041)	(0.253)	(0.005)	(0.010)	(0.009)	(0.009)	(0.026)	(0.034)	(0.026)	(0.020)	(0.001)	(0.029)	(0.043)	(0.033)	(0.041)	(0.030)	(0.026)	(0.033)	(0.033)	(0.047)
Rain	0.012	0.200	0.001	-0.001	0.000	0.016	-0.011	-0.002	0.019	-0.017	-0.004	0.039	-0.008	-0.027	-0.003	-0.015	0.044*	-0.007	0.002	0.030
	(0.032)	(0.217)	(0.003)	(0.003)	(0.002)	(0.014)	(0.022)	(0.029)	(0.025)	(0.013)	(0.003)	(0.025)	(0.025)	(0.029)	(0.027)	(0.028)	(0.025)	(0.027)	(0.024)	(0.049)
Mean	0.53	113.75	0.00	0.99	0.00	0.04	0.20	0.40	0.24	0.10	0.00	0.24	0.25	0.24	0.25	0.43	0.23	0.58	0.63	1.45

	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)	(31)	(32)	(34)	(35)	(36)	(36)	(37)	(38)	(39)
	No. female adult	No. male child	No. female child	No. female teen	No. male teen	No. persons	Consumption index	Village electricity	Village phone %	Village road pucca	Village road katcha	Village road no	Village Immunisation	Village health checkups	Village nutrition	Village growth monitor	Village early child	Village girls adoles	Village no. immu.
TSC	-0.091**	-0.038	0.078	-0.028	-0.006	-0.167	0.008	0.068	-1.312	0.015	-0.004	-0.011	-0.010	-0.020	0.003	-0.004	-0.019	-0.044**	-0.267
	(0.045)	(0.053)	(0.082)	(0.030)	(0.027)	(0.128)	(0.038)	(0.241)	(0.812)	(0.017)	(0.015)	(0.009)	(0.016)	(0.022)	(0.015)	(0.016)	(0.019)	(0.019)	(0.174)
Rain	0.013	0.035	0.056	-0.017	0.023	0.140	-0.019	-0.150	1.065	0.007	-0.010	0.003	0.015	0.001	-0.004	0.023	0.003	-0.006	0.180
	(0.046)	(0.056)	(0.059)	(0.032)	(0.027)	(0.135)	(0.030)	(0.167)	(0.867)	(0.016)	(0.017)	(0.006)	(0.011)	(0.018)	(0.016)	(0.015)	(0.015)	(0.018)	(0.195)
Mean	1.52	1.53	1.41	0.24	0.16	6.34	9.61	12.96	84.63	0.88	0.10	0.01	0.93	0.84	0.91	0.85	0.85	0.48	5.10

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The sample is the same as the ones used in our main analysis. Each estimate presents a coefficient and a clustered standard error in parentheses from a separate regression of the variables listed as the dependent variable, on the TSC intensity and rain, with district, birth year, year of survey and state×birth year fixed effects.